CCSS Advanced Algebra

Supplement to CPM *Algebra 2 Connections*

Portland Public Schools
CCSS Advanced Algebra Implemented 2014-2015

Algebra 2 Connections

CCSS Supplemental Materials

Student Edition, Version 1.1

Managing Editor

Bob Petersen
Sacramento, CA

Contributing Editors

Scott Coyner
Sacramento, CA

Leslie Dietiker
Michigan State University

Lori Hamada
Fresno, CA

Brian Hoey
Sacramento, CA

Michael Kassarjian
Dallas, TX

Judy Kysh
San Francisco, CA

Sarah Maile
Sacramento, CA

Technical Manager

Sarah Maile
Sacramento, CA

Program Directors

Leslie Dietiker
Michigan State University
East Lansing, MI

Brian Hoey
Christian Brothers High School
Sacramento, CA

Judy Kysh, Ph.D.
Departments of Mathematics and Education
San Francisco State University

Tom Sallee, Ph.D.
Department of Mathematics
University of California, Davis

1 2 3 4 5 6 7 8 9 15 14 13 12 11

Printed in the United States of America ISBN 978-1-60328-056-3

Algebra 2 Connections
CCSS Supplemental Materials

OPERATIONS WITH RATIONAL EXPRESSIONS

10.1.1 How can I simplify?

Simplifying Expressions

In Chapter 8, you used the special qualities of the number zero to develop a powerful way to solve factorable quadratics. In Section 10.1, you will focus on another important number: the number 1. What is special about 1? What can you do with the number 1 that you cannot do with any other number? You will use your understanding of the number 1 to simplify algebraic fractions, which are also known as **rational expressions**.

10-1. What do you know about the number 1? Brainstorm with your team and be ready to report your ideas to the class. Create examples to help show what you mean.

10-2. Mr. Wonder claims that anything divided by itself equals 1 (as long as you do not divide by zero). For example, he says that $\frac{16x}{16x}=1$ if x is not zero.

a. Is Mr. Wonder correct?

b. Why can't x be zero?

c. Next he considers $\frac{x-3}{x-3}$. Does this equal 1? What value of x must be excluded in this fraction?

d. Create your own rational expression (algebraic fraction) that equals 1. **Justify** that it equals 1.

e. Mr. Wonder also says that when you multiply any number by 1, the number stays the same. For example, he says that the product below equals $\frac{x}{y}$. Is he correct?

$$\frac{z}{z}\cdot\frac{x}{y}=\frac{x}{y}$$

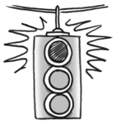
10-3. Use what you know about the number 1 to simplify each expression below, if possible. State any values of the variables that would make the denominator zero.

a. $\frac{x^2}{x^2}$
b. $\frac{x}{x}\cdot\frac{x}{x}\cdot\frac{x}{3}$
c. $\frac{x-2}{x-2}\cdot\frac{x+5}{x-1}$
d. $\frac{9}{x}\cdot\frac{x}{9}$
e. $\frac{h\cdot h\cdot k}{h}$
f. $\frac{(2m-5)(m+6)}{(m+6)(3m+1)}$
g. $\frac{6(n-2)^2}{3(n-2)}$
h. $\frac{3-2x}{(4x-1)(3-2x)}$

10-4. Mr. Wonder now tries to simplify $\frac{4x}{x}$ and $\frac{4+x}{x}$.

a. Mr. Wonder thinks that since $\frac{x}{x}=1$, then $\frac{4x}{x}=4$. Is he correct? Substitute three values of x to **justify** your answer.

b. He also wonders if $\frac{4+x}{x}=5$. Is this simplification correct? Substitute three values of x to **justify** your answer. Remember that $\frac{4+x}{x}$ is the same as $(4+x)\div x$.



c. Compare the results of parts (a) and (b). When can a rational expression be simplified in this manner?

d. Which of the following expressions below is simplified correctly? Explain how you know.

i. $\frac{x^2+x+3}{x+3}=x^2$

ii. $\frac{(x+2)(x+3)}{x+3}=x+2$

10-5. In problem 10-4, you may have noticed that the numerator and denominator of an algebraic fraction must both be written as a product before any terms create a 1. Examine the expressions below. Factor the numerator and denominator of each fraction, if necessary. That is, rewrite each one as a product. Then look for “ones” and simplify. For each expression, assume the denominator is not zero.

a. $\frac{x^2+6x+9}{x^2-9}$

b. $\frac{2x^2-x-10}{3x^2+7x+2}$

c. $\frac{28x^2-x-15}{28x^2-x-15}$

d. $\frac{x^2+4x}{2x+8}$

10-6. In your Learning Log, explain how to simplify rational expressions such as those in problem 10-5. Be sure to include an example. Title this entry “Simplifying Rational Expressions” and include today’s date.

MATH NOTES

LOOKING DEEPER

Multiplicative Identity Property

When any number is multiplied by 1, its value stays the same.

For example:

$142 \cdot 1 = 142$ $\qquad$ $1 \cdot k^2 = k^2$ $\qquad$ $\boxed{\frac{4}{4}} \cdot \frac{2}{3} = \frac{2}{3}$

10.1.2 How can I rewrite it?

Multiplying and Dividing Rational Expressions

In a previous course you learned how to multiply and divide fractions. But what if the fractions have variables in them? (That is, what if they are rational expressions?) Is the process the same? Today you will learn how to multiply and divide rational expressions and will continue to practice simplifying rational expressions.

10-13. Review what you learned yesterday as you simplify the rational expression below. What are the excluded values of x? (That is, what values can x not be?)

$$\frac{3x^2+11x-4}{2x^2+11x+12}$$

10-14. With your team, review your responses to homework problem 10-11. Verify that everyone obtained the same answers and be prepared to share with the class how you multiplied and divided the fractions below.

$\frac{2}{3}\cdot\frac{9}{14}$ $\qquad$ $\frac{3}{5}\div\frac{12}{25}$

10-15. Use your understanding of multiplying and dividing fractions to rewrite the expressions below. Then look for "ones" and simplify. For each rational expression, also state any values of the variables that would make the denominator zero.

a. $\frac{4x+3}{x-5}\cdot\frac{x-5}{x+3}$

b. $\frac{x+2}{9x-1}\div\frac{2x+1}{9x-1}$

c. $\frac{2m+3}{3m-2}\cdot\frac{7+4m}{3+2m}$

d. $\frac{(y-2)^3}{3y}\cdot\frac{y+5}{(y+2)(y-2)}$

e. $\frac{15x^3}{3y}\div\frac{10x^2y}{4y^2}$

f. $\frac{(5x-2)(3x+1)}{(2x-3)^2}\div\frac{(5x-2)(x-4)}{(x-4)(2x-3)}$

10-16. PUTTING IT ALL TOGETHER

Multiply or divide the expressions below. Leave your answers as simplified as possible. For each rational expression, assume the denominator is not zero.

a. $\frac{20}{22} \cdot \frac{14}{35}$

b. $\frac{12}{40} \div \frac{15}{6}$

c. $\frac{5x-15}{3x^2+10x-8} \div \frac{x^2+x-12}{3x^2-8x+4}$

d. $\frac{12x-18}{x^2-2x-15} \cdot \frac{x^2-x-12}{3x^2-9x-12}$

e. $\frac{5x^2+34x-7}{10x} \cdot \frac{5x}{x^2+4x-21}$

f. $\frac{2x^2+x-10}{x^2+2x-8} \div \frac{4x^2+20x+25}{x+4}$

10-17. In your Learning Log, explain how to multiply and divide rational expressions. Be sure to include an example of each. Title this entry “Multiplying and Dividing Rational Expressions’ and include today’s date.

MATH NOTES

METHODS AND MEANINGS

Rewriting Rational Expressions

To simplify a rational expression, both the numerator and denominator must be written in factored form. Then look for factors that make 1 and simplify. Study Examples 1 and 2 below.

Example 1: $\frac{x^2+5x+4}{x^2+x-12} = \frac{(x+4)(x+1)}{(x+4)(x-3)} = 1 \cdot \frac{x+1}{x-3} = \frac{x+1}{x-3}$ for $x \neq -4$ or 3

Example 2: $\frac{2x-7}{2x^2+3x-35} = \frac{(2x-7)(1)}{(2x-7)(x+5)} = 1 \cdot \frac{1}{x+5} = \frac{1}{x+5}$ for $x \neq -5$ or $\frac{7}{2}$

Just as you can multiply and divide fractions, you can multiply and divide rational expressions.

Example 3: Multiply $\frac{x^2+6x}{(x+6)^2} \cdot \frac{x^2+7x+6}{x^2-1}$ and simplify for $x \neq -6$ or 1.

After factoring, this expression becomes: $\frac{x(x+6)}{(x+6)(x+6)} \cdot \frac{(x+1)(x+6)}{(x+1)(x-1)}$

After multiplying, reorder the factors: $\frac{(x+6)}{(x+6)} \cdot \frac{(x+6)}{(x+6)} \cdot \frac{x}{(x-1)} \cdot \frac{(x+1)}{(x+1)}$

Since $\frac{(x+6)}{(x+6)} = 1$ and $\frac{(x+1)}{(x+1)} = 1$, simplify: $1 \cdot 1 \cdot \frac{x}{(x-1)} \cdot 1 \Rightarrow \frac{x}{(x-1)}$

Example 4: Divide $\frac{x^2-4x-5}{x^2-4x+4} \div \frac{x^2-2x-15}{x^2+4x-12}$ and simplify for $x \neq 2, 5, -3$, or -6.

First, change to a multiplication expression: $\frac{x^2-4x-5}{x^2-4x+4} \cdot \frac{x^2+4x-12}{x^2-2x-15}$

Then factor each expression: $\frac{(x-5)(x+1)}{(x-2)(x-2)} \cdot \frac{(x-2)(x+6)}{(x-5)(x+3)}$

After multiplying, reorder the factors: $\frac{(x-5)}{(x-5)} \cdot \frac{(x-2)}{(x-2)} \cdot \frac{(x+1)}{(x-2)} \cdot \frac{(x+6)}{(x+3)}$

Since $\frac{(x-5)}{(x-5)} = 1$ and $\frac{(x-2)}{(x-2)} = 1$, simplify to get:

$\frac{(x+1)(x+6)}{(x-2)(x+3)} \Rightarrow \frac{x^2+7x+6}{x^2+x-6}$

Note: From this point forward in the course, you may assume that all values of x that would make a denominator zero are excluded.

12.1.2 How can I rewrite it?

Adding and Subtracting Rational Expressions

So far in this course you have learned a lot about rational expressions. You have learned how to simplify complex algebraic fractions by factoring the numerators and denominators. You have also learned how to multiply and divide rational expressions. What else is there to learn? Today you will develop a method to add and subtract algebraic fractions.

12-11. With your team, review your responses for homework problem 12-10. Verify that everyone obtained the same answers and be prepared to share how you added fractions with the class.

$\frac{8}{11}-\frac{3}{11}$ $\qquad$ $\frac{x}{6}+\frac{2}{6}$ $\qquad$ $\frac{1}{3}+\frac{2}{5}$

12-12. Examine each expression below. For each one:

- Use your understanding of adding fractions to add the algebraic expressions.
- Simplify your solutions, if possible.

a. $\frac{2x}{2x^2+x-21}+\frac{7}{2x^2+x-21}$

b. $\frac{5x}{x^2-2x-3}-\frac{15}{x^2-2x-3}$

c. $\frac{3x+9}{8x^2-50}-\frac{x+4}{8x^2-50}$

d. $\frac{x^2+5x-2}{3x^2+2x-8}+\frac{2x^2-3x-6}{3x^2+2x-8}$

12-13. What if the algebraic fractions do not have the same denominator? With your team, discuss how to add the fractions below. Be prepared to **justify** your strategy with the class.

a. $\frac{x}{3x+1}+\frac{2x^2-2}{(x-5)(3x+1)}$

b. $\frac{9-3x}{(x+3)(x-3)}+\frac{2x}{x+3}$

12-14. Estacia wants to learn more about excluded values.

a. Explain to Estacia why x cannot be 4 in the expression $\frac{x+2}{x-4}$.

b. Find the excluded values of x in each of the expressions of problem 12-13.

c. With your team, create an expression that has the excluded values of $x \neq -6$ and $x \neq \frac{1}{3}$. Be prepared to share your expression to the class.

12-15. In your Learning Log, explain how to add and subtract rational expressions. Be sure to include an example. Title this entry "Adding and Subtracting Rational Expressions" and include today's date.

12.1.3 How can I rewrite it?

More Adding and Subtracting Rational Expressions

Today you will complete your work with rational expressions. By the end of this lesson you will know how to add, subtract, multiply, and divide rational expressions.

12-22. Review what you learned in Lesson 12.1.2 by adding and subtracting the expressions below. Leave your solutions as simplified as possible.

a. $\frac{5}{8}+\frac{1}{6}$

b. $\frac{8}{9}-\frac{2}{3}$

c. $\frac{x+5}{x+2}+\frac{2x+1}{x+2}$

d. $\frac{x^2-3}{(x+5)(2x-1)}+\frac{x}{2x-1}$

12-23. Examine the expression below.

$$\frac{2x-1}{3x^2+13x+4}+\frac{x+3}{x^2-3x-28}$$

a. With your team, decide how you can alter the expression so that the fractions have a common denominator. Be ready to share your idea with the class.

b. If you have not already do so, add the fractions. Then simplify the result, if possible.

c. Repeat the process to subtract the expressions below. Simplify the result, if possible.

$$\frac{2}{x+4}-\frac{4x-x^2}{x^2-16}$$

12-24. PULLING IT ALL TOGETHER

You now know how to add, subtract, multiply, and divide rational expressions. Pull this all together by simplifying the following expressions.

a. $\frac{x^2-3x-10}{x^2-4x-5} \div \frac{x^2-7x-18}{2x^2-5x-7}$

b. $\frac{2x^2+x}{(2x+1)^2}-\frac{3}{2x+1}$

c. $\frac{15x-20}{x-5} \cdot \frac{x^2-2x-15}{3x^2+5x-12}$

d. $\frac{4}{2x+3}+\frac{x^2-x-2}{2x^2+5x+3}$

e. $\frac{6x-4}{3x^2-17x+10}-\frac{1}{x^2-2x-15}$

f. $\frac{x^2-x-2}{4x^2-7x-2} \div \frac{x^2-2x-3}{3x^2-8x-3}$

MATH NOTES

METHODS AND MEANINGS

Adding and Subtracting Rational Expressions

In order to add and subtract fractions, the fractions must have a common denominator. One way to do this is to change each fraction so that the denominator is the **least common multiple** of the denominators. For the example at right, the least common multiple of $(x+3)(x+2)$ and $x+2$ is $(x+3)(x+2)$.

$$\frac{4}{(x+2)(x+3)}+\frac{2x}{x+2}$$

The denominator of the first fraction already is the least common multiple. To get a common denominator in the second fraction, multiply the fraction by $\frac{(x+3)}{(x+3)}$, a form of the number 1.

$$=\frac{4}{(x+2)(x+3)}+\frac{2x}{x+2}\cdot\frac{(x+3)}{(x+3)}$$

Multiply the numerator and denominator of the second term.

$$=\frac{4}{(x+2)(x+3)}+\frac{2x(x+3)}{(x+2)(x+3)}$$

Distribute the numerator, if necessary.

$$=\frac{4}{(x+2)(x+3)}+\frac{2x^2+6x}{(x+2)(x+3)}$$

Add, factor, and simplify the result.

$$=\frac{2x^2+6x+4}{(x+2)(x+3)}=\frac{2(x+1)(x+2)}{(x+2)(x+3)}=\frac{2(x+1)}{(x+3)}$$

THE FUNDAMENTAL THEOREM OF ALGEBRA

The Fundamental Theorem of Algebra states that a polynomial of degree *n* with complex coefficients has *n* roots in the complex numbers. This also means that the polynomial has *n* linear factors since for every root $x = a$, $(x - a)$ is a linear factor.

Note that the total of *n* roots might included roots that are occur multiple times. If the polynomial contains three factors that are the same, the associated root has a **multiplicity** of three.

QUADRATIC POLYNOMIALS

Some quadratic ($n = 2$) polynomials can easily be factored using integers and have rational roots. The roots of quadratics not factorable with integers are commonly found using the quadratic formula and are real or complex numbers. From each of these roots, liner factors can be created.

For each quadratic polynomial, find the roots and the linear factors.

Example 1

$2x^2 - 3x - 5$: Use a generic rectangle or guess and check to factor.

$2x^2 - 3x - 5 = (2x - 5)(x + 1)$ and the roots are $x = \frac{5}{2}, -1$

Example 2

$x^2 - 4x - 1$: Use the quadratic formula to find the roots.

$$x = \frac{-(-4) \pm \sqrt{(-4)^2 - 4 \cdot 1(-1)}}{2 \cdot 1} = \frac{4 \pm \sqrt{20}}{2} = 2 \pm \sqrt{5}$$

The factors are $(x - (2 + \sqrt{5}))(x - (2 - \sqrt{5})) = (x - 2 - \sqrt{5})(x - 2 + \sqrt{5})$

Example 3

$x^2 - 2x + 2$: Use the quadratic formula to find the roots.

$$x = \frac{-(-2) \pm \sqrt{(-2)^2 - 4 \cdot 1(2)}}{2 \cdot 1} = \frac{2 \pm \sqrt{-4}}{2} = 1 \pm i$$

The factors are $(x - (1 + i))(x - (1 - i)) = (x - 1 - i)(x - 1 + i)$

In all cases, multiplying the factors together will yield the original polynomial.

HIGHER DEGREE POLYNOMIALS

The Fundamental Theorem of Algebra states that the roots exist but provides no information about how to find them. If one root can be found by inspection, a graph, or the integral root theorem, it can be used to reduce the polynomial into factors of smaller degree. Other factoring techniques may also be used.

Example

Looking at a graph of $x^3 - 2x^2 + 5x - 24$ shows that $x = 3$ is a root. Find all the roots and the linear factors.

If $x = 3$ is a root then $(x-3)$ is a factor. To find the other factor, divide the original polynomial by $(x-3)$. In this case using generic rectangles:

	x^2	x	8
x	x^3	x^2	$8x$
-3	$-3x^2$	$-3x$	-24

Use the quotient to find the other roots.
For $x^2 + x + 8$ the roots are:

$$x = \frac{-1 \pm \sqrt{1^2 - 4 \cdot 1 \cdot 8}}{2 \cdot 1} = \frac{-1 \pm i\sqrt{31}}{2}$$

The three roots are $x = 3, \frac{-1 \pm i\sqrt{31}}{2}$.

The linear factors are $(x-3)\left(x - \frac{-1 \pm i\sqrt{31}}{2}\right)\left(x - \frac{-1 \pm i\sqrt{31}}{2}\right)$

Problems

Find the roots and the linear factors for each polynomial.

1. $3x^2 + 7x + 2$
2. $x^2 - 2x - 2$
3. $x^2 + 4x + 5$
4. $2x^2 + 5x + 4$
5. $x^2 + 6x + 2$
6. $9x^2 - 12x + 4$
7. $x^2 - 2ix - 5$
8. $ix^2 + 3x + 4i$
9. $x^3 + x^2 - x + 2$ given root $x = -2$
10. $x^3 - 3x^2 + x - 3$ given root $x = 3$
11. $x^3 - 5x^2 + 6x$
12. $x^4 + 1$

For problems 13–16, determine the possible integral roots, find one root and use it to find the other roots and factors.

13. $x^3 - 6x^2 + 11x - 6$
14. $x^3 + 2x^2 - 5x - 10$
15. $x^3 - 3x^2 + x + 5$
16. $x^3 + 8$

REMAINDER THEOREM

For any number c, when a polynomial $P(x)$ is divided by $(x - c)$, the remainder is $P(c)$.

In general, if $P(x)$ is a polynomial of degree n, then $P(x) = Q(x)(x - c) + P(c)$, where $Q(x)$ is a polynomial of degree $n - 1$.

Example 1

Let $P(x) = x^3 - 3x^2 - 7x + 9$. Verify the remainder theorem and find $P(5)$.

Using direct substitution, $P(5) = 5^3 - 3(5)^2 - 7(5) + 9 = 125 - 75 - 35 + 9 = 24$

Solution: Calculate the remainder for $(x^3 - 3x^2 - 7x + 9) \div (x - 5)$.

Using a generic rectangle to complete the division (Lesson 9.3.1):

×				
x				
-5				
x^3	$-3x^2$	$-7x$	9	

⇒

×	x^2	$2x$	3	
x	x^3	$2x^2$	$3x$	$\boxed{24}$
-5	$-5x^2$	$-10x$	-15	
x^3	$-3x^2$	$-7x$	9	

Therefore $(x^3 - 3x^2 - 7x + 9) = (x - 5)(x^2 + 2x + 3) + 24$ and $P(5) = 24$.

Example 2

Let $P(x) = x^4 + 3x^3 - 9x + 7$. Use the remainder theorem to find $P(-2)$.

Solution: Calculate the remainder for $(x^4 + 3x^3 - 9x + 7) \div (x - (-2))$.

Using long division:

$$
\begin{array}{r}
x^3 + x^2 - 2x - 5 \\
x + 2 \overline{)\, x^4 + 3x^3 + 0x^2 - 9x + 7} \\
-\underline{(x^4 + 2x^3)} \qquad\qquad\qquad \\
1x^3 + 0x^2 \qquad\qquad \\
-\underline{(x^3 + 2x^2)} \qquad\qquad \\
-2x^2 - 9x \qquad \\
-\underline{(-2x^2 - 4x)} \qquad \\
-5x + 7 \\
-\underline{(-5x - 10)} \\
\boxed{17}
\end{array}
$$

Therefore $(x^4 + 3x^3 - 9x + 7) = (x + 2)(x^3 + x^2 - 2x - 5) + 17$ and $P(-2) = 17$.

Check: $P(-2) = (-2)^4 + 3(-2)^3 - 9(-2) + 7 = 16 - 24 + 18 + 7 = 17$

Problems

Use the remainder theorem to find $P(c)$ for the given $P(x)$ and number *c*.

1. $P(x) = 2x^3 - 7x^2 + 6x - 3$, $c = 3$

2. $P(x) = x^3 + 6x^2 + 5x - 6$, $c = -2$

3. $P(x) = 2x^4 - 5x^3 - 8x - 3$, $c = -1$

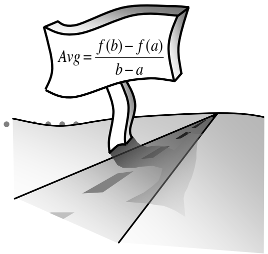
4. $P(x) = 4x^3 - 5x + 6$, $c = 1$

5. $P(x) = 2x^3 - 3x^2 + 2x - 3$, $c = 2$

5. $P(x) = x^4 + 3x^3 + 8x - 2$, $c = -4$

RATES OF CHANGE

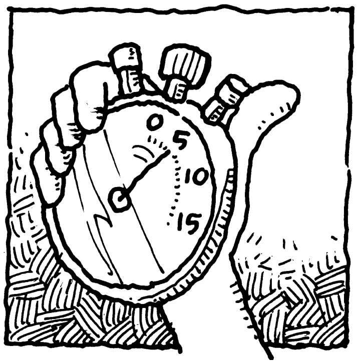
9.1.1 What's my rate?



Rates of Change from Data

Rates occur naturally in many applications. You certainly have experienced how much money you make per hour or how many miles per gallon of gas your car gets, both examples of common rates. Today we will be looking at average rates of change from a variety of data sources.

9-1. We start our introduction to average rate of change with a simple experiment. To begin, you should find your pulse, then count your pulse beats for 15 seconds. You may want to designate a class timer who will call out "start" and 15 seconds later, "stop," during which time you will count your pulse beats. (Do not let the timer begin until you have found your pulse!)



a. Convert your results to beats per minute. This is your **heart rate**.

b. What assumption did you make in order to calculate your heart rate?

c. What activities would make your heart rate change?

9-2. Other rates common in our society include:

Birth rates recorded in births per 1,000 people;

Interest rates recorded in cents per dollar;

ERA's recorded in earned runs per nine innings pitched.

With your team, list at least 3 other rates. Be sure to indicate the units.

9-3. Rates are also used to describe how one quantity is changing in relation to another. This is called a "**rate of change**" or an "**average rate of change**." To illustrate this, consider the following statement: Macario drove from Vacaville to Davis (a distance of 28 miles) in 30 minutes.

a. What was his average speed in miles per hour?

b. Does this mean that he drove that speed the entire trip? If not, what does it mean?

c. Did he *ever* drive the average speed of 56 mph? Discuss this in your team.

9-4. As the Shuttle mission STS-82 lifted off from the NASA launching pad at Cape Canaveral, Mission Control in Houston was receiving the following data about the height of the rocket in feet.

Time (in seconds)	3	6	9	12	15	18	21
Height (in feet)	70	297	701	1306	2131	3200	4533

Find the rocket's average velocity at different times by calculating $\frac{\text{change in distance}}{\text{change in time}} = \frac{\Delta d}{\Delta t}$ for the intervals given in the following table.

Note that the Greek capital letter delta (Δ) in the previous formula means "change in."

a. From $t = 3$ to $t = 21$.

b. From $t = 3$ to $t = 18$.

c. From $t = 3$ to $t = 15$.

d. From $t = 3$ to $t = 12$.

e. From $t = 3$ to $t = 9$.

f. From $t = 3$ to $t = 6$.

9-5. A weather balloon is sent up to measure the air pressure, temperature, and dew point at a particular location. It radios back the following information:

Time (s)	Altitude (ft)	Pressure (mb)	Temp ˚ C	Dew Pt ˚ C
0	ground level	1000	17.8	9
360	2900	925	25.6	4
650	4900	850	22.2	2
1400	9900	700	13	–13
2750	19900	500	–4.1	–26
3600	24900	400	–17.9	–37

a. How fast is the balloon ascending from $t = 360$ s to $t = 650$ s? $\left(\dfrac{\Delta A}{\Delta t}\right)$

Be sure to include the units (ft/s).

b. What is the average rate of change of the pressure with respect to time from $t = 650$ s to $t = 2750$ s? $\left(\dfrac{\Delta P}{\Delta t}\right)$

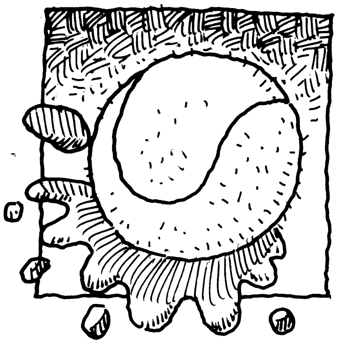
c. What does the negative sign mean in part (b)?

d. How is the temperature changing with respect to the altitude from $t = 0$ s to $t = 3600$ s? $\left(\dfrac{\Delta \text{temp}}{\Delta A}\right)$

Review & Preview

9-6. Another common rate of change we use is slope, $\dfrac{\Delta y}{\Delta x}$. A *linear* function, $f(x)$, includes a point (2, 7) and another point (4, 8). Find the average rate of change, $\dfrac{\Delta y}{\Delta x}$, between these two points. What would happen to the slope if you chose 2 other points on the function?

9-7. A child tosses a tennis ball into the air and it lands in the mud. The function $h(t) = -16(t - 0.5)^2 + 8$ gives the ball's height in feet with respect to time in seconds. (When you study physics or calculus you will learn how to find such functions yourself.)



a. Sketch a graph of the function $h(t)$. Darken the portion of the curve that fits our scenario.

b. What is happening to the height of the tennis ball with respect to the time?

c. Since speed is a function of distance traveled over time, what part of the graph represents the speed of the ball? What is happening to the speed of the tennis ball with respect to time?

d. Fill in the following table of time (s) versus height (ft). Be careful. Remember that h cannot be negative in this situation.

t seconds	0	0.25	0.5	0.75	1	1.25	1.5
h in feet							

e. Find the average velocity, $\frac{\Delta d}{\Delta t}$, in ft/s for each 0.25 second time interval.

f. Does the ball have the same velocity on any of these intervals? If so, which ones? Think carefully, this may be tricky.

9-8. The dew point is the temperature at which the air would be completely saturated with water (100% humidity). How could we use Δ to represent the change in dew point with respect to the change in height? Using the data from the balloon problem, calculate the average rate of change of the dew point with respect to height from $t = 0$ s to $t = 3600$ s.

9-9. A few years ago, Carlos was driving on the New Jersey Turnpike (a toll road that collects money every few miles). He was presented with a speeding ticket as he reached the last booth. Since there were no highway patrolmen or radar on the turnpike, he could not figure out how anyone knew that he had been speeding.

a. How do you suppose his speed had been determined?

b. Using your conclusion in part (a), do you think it would be possible for someone to speed on part of their turnpike trip and not get caught?

9-10. A car which runs very slowly in the morning travels at a speed of $2t$ ft/s in the first t seconds after it starts for $0 \le t \le 50$.

a. Draw the velocity versus time graph.

b. Use your graph to find the distance the car travels in the first t seconds after it starts.

9-11. Suppose you want to buy rice. Different size packages cost different amounts, but the relationship is not always linear. That is, a bag twice as big does not usually cost twice as much. The chart shows the prices for various sizes of bags of rice.

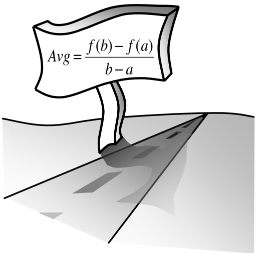
1/2 lb bag	$0.89
1 lb bag	$1.29
2 lb bag	$1.89
5 lb bag	$4.60
10 lb bag	$8.95
20 lb bag	$17.80

a. Find the rates in cost per pound. (Stores refer to this as unit pricing.)

b. Does the unit price increase or decrease with the size of the bag?

c. Does the rate change more drastically for smaller sizes or for larger sizes?

9.1.2 How does it change?



More Rates of Change from Data

In the previous lesson you saw how rates of change could be calculated from data. In today's lesson we will continue to look at data and extend our understanding to the graphs of the data as well. We will look at how the average rate of change can be seen from both data and a graph.

9-17. The table below shows the population of Australia for the given years.

Year	Population (millions)	Year	Population (millions)	Year	Population (millions)
1970	12.51	1987	16.10	1997	18.52
1973	13.27	1989	16.83	1999	18.94
1977	14.07	1991	16.85	2002	19.66
1981	14.57	1992	17.07	2004	20.10
1986	15.61	1996	18.06	2007	20.74

a. Enter the data above into your calculator. Create a plot of the data and find an equation to model this data.

b. If the population growth of Australia continues in this pattern, what population can they expect in the year 2020?

c. From the table, what was the average rate of growth $\left(\frac{\Delta p}{\Delta t}\right)$ between 1970 and 1999?

d. Using your model, what will be the average rate of growth $\left(\frac{\Delta p}{\Delta t}\right)$ between 1999 and 2020?

e. When is the growth increasing most rapidly?

9-18. Banks advertise rates in both the actual rate and the annual percentage rate (APR). Suppose you deposited $1000 into a bank that offers 6% annual interest compounded monthly.

a. How much money would you have after one year?

b. What percent did you earn for the year (this is the APR)?

c. What is the APR if the bank offers continuous compounding?

9-19. When a person wins the “Dreams Come True” Lottery for $1,000,000, they receive $50,000 immediately and $50,000 a year for 19 more years. “Dreams Come True” deposits money into a 12% savings account in order to make the yearly $50,000 payments. At this rate of interest, they only need to deposit $368,288.84 in order to make the payments and have the account “zeroed out” after 19 years.

a. What is the average rate of change over the 19 years that the money is in the account?

b. If the payments are $50,000 per year, why is the balance in the account not dropping by $50,000 a year?

This is a chart of the amount of money in the account at the end of each year.

0	$368,288.84	10	$266,412.49
1	$362,483.50	11	$248,381.99
2	$355,981.52	12	$228,187.83
3	$348,699.31	13	$205,570.37
4	$340,543.22	14	$180,238.81
5	$331,408.41	15	$151,867.47
6	$321,177.42	16	$120,091.56
7	$309,718.71	17	$84,502.55
8	$296,884.96	18	$44,642.86
9	$282,511.15	19	0

c. Enter the data into your calculator to create a statplot showing the amount of money in the account over the 19 year period. Describe the graph of the data.

d. Calculate the actual change of the account each year. (Divide up the work in your team.)

e. Do any of these match the average rate of change for the 19 years?

f. Is the average rate of change per year increasing or decreasing?

g. Describe the concavity of the graph in part (c). Is the graph concave up or concave down? What does that tell you about the slope of the function?

9-21. a. Find the average rate of change for the function $a(x)=3x-5$ over the interval [2, 6].

b. Repeat part (a) for the interval [5, 7].

c. Do you think it is true that $a(x)$ will have the same average rate of change over *any* interval? Why or why not?

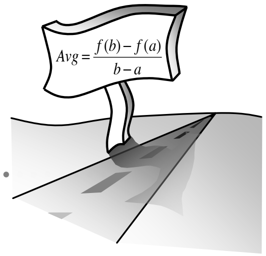
9-22. Which (if any) of the following would have a fairly constant rate of change? Explain your answer.

a. The speed of a car as it enters an on-ramp to a freeway until it reaches the freeway speed limit.

b. The rise in temperature of an oven going from room temperature to 350°.

c. The number of kilowatt-hours of electricity used by one household from 7:00 AM to 10:00 AM.

9.1.3 How do I find rates of change for functions?



Slopes and Rates of Change

In the previous lesson, you looked at rates of change from data. But what about rates of change for a function? How can we calculate average rates of change directly from the function? In this lesson we look at how functions change.

9-31. A toy car travels a distance $10t^2$ feet in the first t seconds after it starts. Calculate the average velocity of the car on these intervals by calculating the change in distance divided by the change in time.

a. $[3, 4]$

b. $[3, 3.1]$

c. $[3, 3.01]$

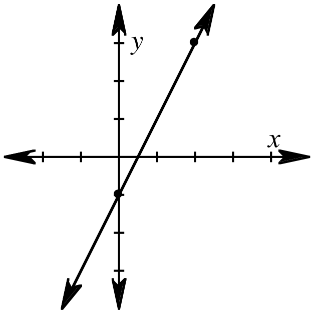
d. $[3, 3.001]$

e. $[2.99, 3]$

f. $[2.999, 3]$

9-32. On the basis of the calculations made in the previous problem, what do you think is the *exact* velocity of the car 3 seconds after it starts?

9-33. Look at the graph to the right. It goes through the points $(0, -1)$ and $(2, 3)$.



a. Explain how you know its equation is $y = 2x - 1$.

b. What is the average rate of change of y with respect to x between $x = -1$ and $x = 1$?

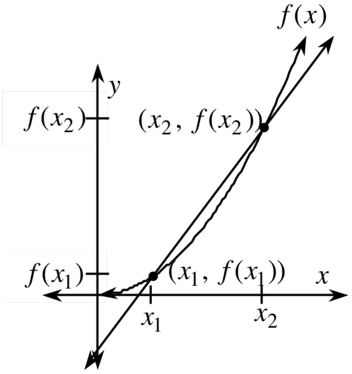
c. What is the average rate of change between $x = 1$ and $x = 5$?

d. What is the average rate of change between $x = 2$ and $x = a$ (a = any constant)?

e. What do you notice about the average rate of change for a linear function?

f. How is the average rate of change related to the equation itself?

9-34. Can we represent an average rate of change graphically? The picture at right is of a curved function, $f(x)$, that passes through the points $(0,0)$, $(x_1, f(x_1))$, and $(x_2, f(x_2))$. You can also see a line that has been drawn through two points on $f(x)$. Use the graph at right to answer the following questions.



a. Write an expression for Δx, the change in x from x_1 to x_2.

b. Write an expression for $\Delta f(x)$, the change of the outputs of the function over that same interval?

c. The average rate of change for the function on the interval $\left[x_1, x_2\right]$ is the same as the slope of the line through corresponding points on the curve: $(x_1, f(x_1))$ and $(x_2, f(x_2))$. Write an expression for this average rate of change.

9-35. Consider the equation $y = e^x$.

a. Sketch the graph and find the average rate of change over the following intervals: $[0, 0.1]$, $[1, 1.1]$, $[2, 2.1]$.

b. What do you notice about the average rate of change for this exponential function?

c. From looking at the graph, for what values of x does the function appear to have the smallest average rate of change in the interval $\left[x, x+0.1\right]$? When does the change appear to be the greatest?

9-36. For each of the following functions, is the average rate of change over the given interval positive or negative? Try to do this without substituting the values.

a. $f(x) = x^2$ over the interval $[2, 3]$.

b. $g(x) = \log x$ over the interval $[0.01, 1]$.

c. $h(x) = 7 - 3x^2$ over the interval $[2, 5]$.

d. $j(x) = 0.5^x$ over the interval $[-1, 0]$.

e. $k(x) = 1.5^x$ over the interval $[-1, 0]$.

f. $m(x) = x^2$ over the interval $[-2, 3]$.

9-37. Now we want to calculate the average velocity of our toy car (with $d(t) = 10t^2$). Find the average velocity on each of the intervals.

a. $[4, 5]$

b. $[4, 4.1]$

c. $[4, 4.01]$

c. $[3.99, 4]$

e. $[3.999, 4]$

9-40. Find the equations of the following lines.

a. The line with slope = $-\frac{1}{2}$ through (5, –2).

b. The line through the points (1, 4) and (3, 0).

9-42.

a. What is the average rate of change of the function $g(x) = 6 - 2x$ over the interval [2, 6]?

b. Over the interval [5, 7]?

c. Do you think it is true that $g(x)$ will have a constant average rate of change over *any* interval? Why or why not?

d. Prove your answer to part (c) by computing the average rate of change for $g(x)$ over the interval $[a, b]$.

9-45. Given $f(x) = 10x^2$ and $g(x) = 15 - 3x$, find the following:

a. $f(3+h) - 90$

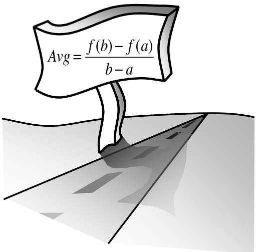
b. $f(x+h) - 10x^2$

c. $g(20+h)$

d. $f\left(g(x)\right)$

9-46. Find the slope between the points (3, 9) and $(3+h,\ 9+6h+h^2)$.

9.1.4 How fast am I going?

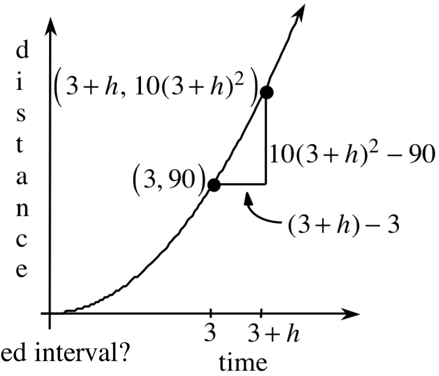


Average Velocity and Rates of Change

Over the last few lessons you have calculated the average rate of change of a function between two points by finding the slope of a line connecting those two points. In this lesson you will investigate what happens when we look at the average rate of change between two points on a function and then allow those two points to get closer and closer together. You will also explore what this value tells you about the function.

9-47. Let $d(t) = 10t^2$ be a distance function and find the average velocity from $t = 5$ to $t = 5.5$. Find the average velocity from $t = 5$ to $t = 5.1$ and then from $t = 5$ to $t = 5.01$.

9-48. We want to generalize the procedure you used in the last problem to any function and any interval. We will begin by computing the average velocity of our toy car ($d(t) = 10t^2$) from $t = 3$ to $t = 3 + h$. As always, we compute the average velocity by dividing the change in distance (Δd) by the change in time (Δt).



a. What is the change in time over the specified interval?

b. Show that the change in distance over this time interval is $60h + 10h^2$.

c. What is the average velocity over this time interval?

9-49. Use the same distance function as in the previous problem: $d(t) = 10t^2$.

a. Compute the average velocity from $t = 4$ to $t = 4 + h$.

b. Use the expression from part (a) to calculate the average velocity from $t = 4$ to $t = 4.1$, $t = 4$ to $t = 4.01$, $t = 3.99$ to $t = 4$, and $t = 3.999$ to $t = 4$.

9-50. In problem 9-20, the volume of the tube was $V = x^2(12-4x)$ with a maximum at $x = 2$. Use the function to find the slope over the following intervals:

a. $[2, 2.5]$

b. $[2, 2.1]$

c. $[2, 2.01]$

d. $[2, 2.001]$

e. $[1.99, 2]$

f $[1.999, 2]$

g. What value is the slope approaching as Δx gets smaller?

9-51. Let $R(t)$ be the altitude of a rocket t seconds after takeoff.

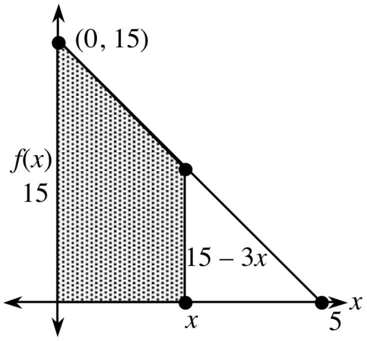
a. Write the expression for the average velocity of the rocket from time $t = 20$ to time $t = 20 + h$ in terms of $R(t)$.

b. As h gets closer and closer to 0, what happens to the average velocity of the rocket?

9-52. Consider the graph of $f(x) = 15 - 3x$ from $0 \le x \le 5$ at the right.

a. Give a general expression for the area under the curve for the interval $[0, x]$, where $0 \le x \le 5$. Hint: Use the area formula for a trapezoid.

b. Give a general expression for the area under the curve for the interval $[1, x]$ where $1 \le x \le 5$.



9-53. Answer the following questions about average rates of change.

a. For a general function $f(x)$, what does $\dfrac{f(5)-f(3)}{5-3}$ represent?

b. If $f(x) = x^2 + 3x + 2$, find $\dfrac{f(5)-f(3)}{5-3}$.

9-54. Given the function $f(x) = x^2 + 3x + 2$, show that the average rate of change between the points $x = 4$ and $x = 4 + h$ is $11 + h$.

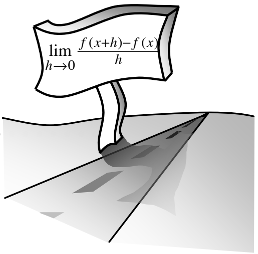
9-55. What is the average rate of change for the following functions over the given intervals?

a. $f(x) = \sin x \quad \left[\dfrac{\pi}{2}, \pi\right]$

b. $f(x) = \log x \quad [1, 10]$

9.2.1 Can I find the slope at a point?

Moving From AROC to IROC



In the past few lessons you have done several problems where you investigated the average rate of change for functions over different intervals. In this lesson, you will learn how to find the rate of change at a specific point.

9-61. In previous lessons you investigated the average rates of change for several linear functions over various intervals. You found that the average rate of change for $y = 3x - 5$ was 3, the average rate of change of $y = 2x - 1$ was 2, and the average rate of change of $y = 6 - 2x$ was -2.

a. As a team, write a general statement about the average rate of change of linear functions.

b. How sure are you of your statements? Have they been proven or are you working from patterns?

9-62. To prove the statement you made in the previous problem, let $f(x) = mx + b$. We want to find the average rate of change of $f(x)$ over the interval $[c, d]$.

a. Explain why we want to compute the ratio $\frac{f(d) - f(c)}{d - c}$.

b. Compute the ratio from part (a) and show that it equals m, the slope.

c. Explain why what you have done shows that the average rate of change of *every* linear function is constant over *every* interval.

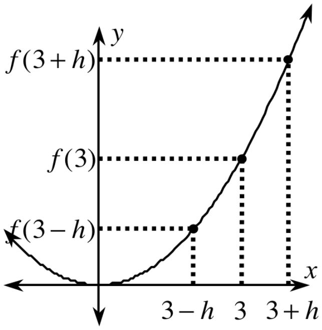
9-63. Now investigate what happens when we look at the average rates of change for a non-linear function. Recall that the average rate of change over an interval was defined as the slope of a line connecting the endpoints of the function values for that interval. (see problem 9-34 (c) if you need help) Let $f(x) = x^2$. Find the slope of:

a. The line joining $f(3)$ and $f(4)$ or … (3, 9) and (4, 16).

b. The line joining $f(3)$ and $f(3.1)$ or … (3, 9) and $(3.1, 3.1^2)$.

c. The line joining $f(3)$ and $f(2.99)$ or … (3, 9) and $(2.99, 2.99^2)$.

9-64. Can we generalize from the patterns of the previous problem to find the slope of the line joining (3, 9) and any other *nearby* point on the curve $f(x) = x^2$?



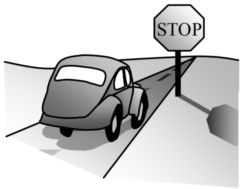
a. Find the average rate of change (the slope), between the two points 3 and $3+h$. Simplify your answer as much as possible.

b. Find the average rate of change between the two points 3 and $3-h$. Again, simplify your answer as much as possible.

c. How close to 3 can you get? In (a) and (b) we were h units away from 3. Can we get closer? Explain.

d. Compare your answers to parts (a) and (b). What happens to the average rate of change as we get closer and closer to $x = 3$? In other words, what happens to the slope as $h \to 0$?

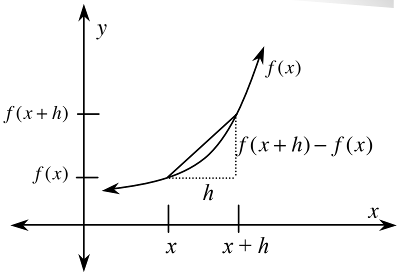
MATH NOTES



AROC and IROC

The **average rate of change** for $f(x)$ in the interval $[x, x+h]$ is:

$$\text{AROC} = \frac{f(x+h)-f(x)}{(x+h)-x} = \frac{f(x+h)-f(x)}{h}$$



The **instantaneous rate of change** for $f(x)$ in the interval $[x, x+h]$ is:

$$\text{IROC} = \lim_{h\to 0} \frac{f(x+h)-f(x)}{(x+h)-x} = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

9-65. Consider the function $r(t) = 2t^3$.

a. Show that the average rate of change of $r(t)$ from $t = 2$ to $t = 2 + h$ is $24 + 12h + 2h^2$.

b. If you wish to graph the function for the average rate of change, what letter do you use on your calculator to represent $r(t)$? What letter could you use to represent h?

c. Write the exact expression you would enter into your calculator in order to graph your answer to part (a).

d. We want to look at what happens as $h \to 0$ (i.e. $x \to 0$). Use your TRACE button to estimate the limit of the average rate of change as $h \to 0$. This value is the IROC or instantaneous rate of change when $t = 2$.

9-66. Given the function $f(x) = 2^x$.

a. Find the average rate of change of $f(x)$ from $x = 5$ to $x = 5 + h$.

b. Use the [TABLE] function on your calculator to evaluate your expression from part (a) when $h = 0$. Why doesn't it give you a value?

c. Graph your expression from part (a). (You will be using x for h in order to do this, just as you did in the previous problem.)

d. Use TRACE to estimate the value of the expression as $h \to 0$. What does this value represent?

9-67. a. Simplify $\frac{f(5+h)-f(5)}{(5+h)-5}$ for $f(x) = x^2 + 2x - 1$.

b. Explain what your answer represents on the graph of $y = f(x)$.

9-69. Let $g(x) = \frac{10}{x}$.

a. Find the average rate of change of $g(x)$ from $x = 5$ to $x = 6$.

b. Write the expression for the average rate of change of $g(x)$ from $x = 5$ to $x = 5 + h$.

c. Use your calculator to evaluate the expression in (b) as $h \to 0$.

9-70. Let $s(x) = \sin x$.

a. Find the exact value of the average rate of change of $s(x)$ from $x = \frac{\pi}{4}$ to $x = \frac{\pi}{2}$.

b. Write the expression for the average rate of change of $s(x)$ from $x = \frac{\pi}{4}$ to $x = \frac{\pi}{4} + h$.

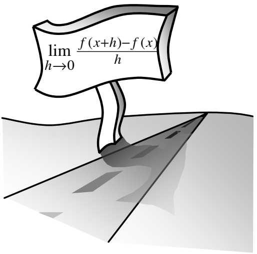
c. Use the fact that $\sin(A+B) = \sin A \cos B + \cos A \sin B$ to show that the expression in part (b) can be written as $\frac{1}{\sqrt{2}}\left(\frac{\cos(h) + \sin(h) - 1}{h}\right)$.

9-71. Write a simplified expression for the slope of the line that goes through the points $(x, f(x))$ and $(x+h, f(x+h))$. The answer will be an algebraic expression, not a number. This expression should look familiar. What is another name for this expression?

9-73. Find the equation of the line tangent to the circle $x^2 + y^2 = 25$ at the point (3, 4).

9-74. Find the equation of the line connecting the points on the function $y = \sqrt{x+1}$ at $x = 3$ and $x = 8$.

9.2.2 What is the slope of a tangent line?



Secant and Tangent Lines

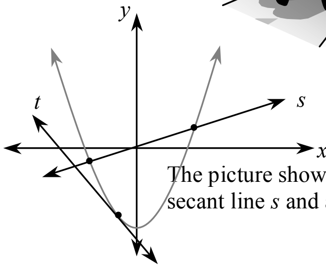
You have seen how slope, the average rate of change, and the instantaneous rate of change are all related. This lesson looks specifically at the lines that we have been using to calculate these two types of rates of change. We begin by developing a general idea of what we mean by secant and tangent lines as they relate to graphs.

MATH NOTES

Secant and Tangent Lines

A **secant line** of a curve is a line that passes through any two distinct points on the curve.

A **tangent line** is a line that "grazes" a curve at a single point.



The picture shows examples of a secant line *s* and a tangent line *t*.

9-75. Obtain the resource page from your teacher for today's lesson. The first graph on the page is the function $f(x) = -x^2 + 4x + 1$.

a. The secant line through $(1, f(1))$ and $(3, f(3))$, is drawn on the graph. Calculate the slope and the average rate of change between these two points.

b. Draw a second secant line through $(1, f(1))$ and $(2, f(2))$ on the graph. Calculate the slope and the average rate of change between these two points.

c. Draw a third secant line through $(1, f(1))$ and $(1.5, f(1.5))$ on the graph. Calculate the slope and the average rate of change between these two points.

d. In the three cases above, you should notice that the secant line is passing through $(1, f(1))$ and points which are closer and closer to $(1, f(1))$. As the points get closer to each other, what happens to the slope of the secant line?

9-76. Use your results from the previous problem to answer the following questions.

a. What can you conclude about the average rate of change of the function $f(x)$ between two points and the slope of the secant line passing through those two points on the graph?

b. Explain why the secant line determined by two points gets closer and closer to a tangent line to the curve as the two points get closer and closer together.

9-77. The second graph on the resource page is $g(x)=\sqrt{3x+1}$. Use it to complete the following problems.

a. Draw the secant line on the graph for $x=1$ and $x=5$.

b. Calculate the slope of the secant line between $x=1$ and $x=5$.

c. Calculate the average rate of change between these two points.

d. Repeat parts (b) and (c) for $x=1$ and $x=3$.

e. Repeat parts (b) and (c) for $x=1$ and $x=2$.

9-78. a. Write the algebraic expression for the slope of $g(x)$ in problem 9-77 of the secant line between $x=1$ and $x=1+h$.

b. Evaluate your expression for slope when $h=0$.

c. Use your calculator to find the slope of the secant line through $(1, g(1))$ and $(1+h, g(1+h))$ when h is very close to 0.

9-79. Consider the function $f(x)=2x^2+5$.

a. Write an expression for the average rate of change from $x=1-h$ to $x=1$.

b. Graph the expression you wrote as a function of h.

c. This average rate of change function approaches a number as $h \to 0$. What is the number?

d. Geometrically, what is the significance of this number?

9-80. If the average rate of change between two points of a function is $\frac{3}{2}$, what can be said about the secant line connecting those two points?

9-81. Mathematics is full of all kinds of amazing equations, but it is comforting to know that some things remain constant for all of them. For example, to find the slope of the tangent line of a function at $x = 2$, you can follow the same procedure, regardless of the function. For each of the following functions, first find the formula for the slope of a secant line for the function from 2 to $2 + h$, then let $h \to 0$ to find the slope of the tangent line at $x = 2$. Use your calculator to estimate the slopes to three decimal places. Enter a small value for h (such as $h = 0.001$) in the slope formula to approximate your answer.

a. $f(x) = 2x^2$

b. $f(x) = 3^x$

c. $f(x) = \log x$

d. $f(x) = \cos x$ (x in radians)

9-82. Use your slopes from problem 9-81 to write the equation of the line which is tangent to each of the functions at the point when $x = 2$. Use point-slope form.

9-83. Describe how your procedure would have been different in the previous problem if you wanted to find the slope of the tangent line at $x = 5$.

9-88. Anne began her commute home by driving at a constant rate of 30 mph for 15 minutes. She then got on the freeway where she drove at a constant rate of 60 mph for 30 minutes. She exited the freeway and immediately had to stop and wait for 15 minutes due to road construction. She then completed her drive home by driving at 30 mph for 15 minutes. Her total commute time was one hour fifteen minutes. What was the distance of Anne's commute? What was her average speed?

9-89. Sketch two graphs, one to show Anne's rate over time and one to show the distance she traveled over time.

NORMAL DISTRIBUTIONS

5.2.3 How can I compare distributions?

Standard Normal Distribution

5-59. In addition to using your calculator to calculate statistics, you can also use it to display data in the form of a histogram or other pictorial display. The following data is a set of scores from all of the applicants for a police academy. Enter the data below into your calculator.

84	73	55	93	85	78	66	97	89	75
94	80	77	78	45	68	90	62	85	100
98	71	82	87	94	75	87	82	80	64

checksum = 2394

a. The president of the academy wants to analyze data on the latest applicants. Calculate the mean and standard deviation for the scores. Justify your choice of standard deviation.

b. Use your calculator to create a histogram, box plot, and five-number summary of the data above. Sketch the displays on your paper and describe the center, shape, and spread of the data.

c. The academy will automatically accept anyone who scores at least one standard deviation above the mean. Which scores will gain automatic acceptance? How many candidates are accepted?

5-60. In the previous problem, the candidates who are accepted have a ***z*-score** of at least 1 since their scores are more than one standard deviation above the mean. A *z*-score is used to standardize a **raw score** (the actual original score). This allows statisticians to compare scores from different individuals who may have taken a different test. If a *z*-score of 1 indicates that the corresponding raw score was 1 standard deviation above the mean:

a. What does a *z*-score of 0 indicate?

b. What does a *z*-score of –1 indicate?

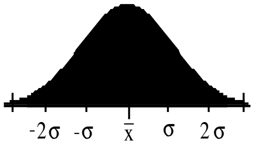
5-61. Joren scored an 82 on his police academy test. He wants to know his z-score so he can figure out if he will be accepted.

a. Use the mean and standard deviation you found in part (a) of problem 5-59 to determine Joren's z-score.

b. Explain to Joren how he can use mean and standard deviation to calculate his z-score. Then use statistics variables to write a formula to calculate z-scores.

5-62. Ryan finds out that applicants for the academy can take a second test if the z-score from the original test is greater than –0.5. Ryan had a raw score of 73. Does Ryan qualify for the second test?

5-63. The diagram at right shows the way data becomes distributed about the mean. It is an important theorem in statistics that averages eventually look like a **normal curve**. In general, we like to measure distance from the mean in "standard deviation units." In a normal distribution, 68% of the population is within approximately 1 standard deviation (1σ) of the mean (μ) while 95% of the population is within 2 standard deviations (2σ) of the mean.



a. One of the most common places where the normal curve is used is on the SAT tests given by the College Board. They scale the raw scores for the verbal and math sections so that the scaled mean will be 500 with a standard deviation of 100. Sketch and label a normal curve, like the one above, that represents the distribution for a scaled math SAT.

b. Mary scored 550 on her math SAT. Mark the location on your sketch from part (a).

c. What was Mary's z-score on the SAT?

d. As stated above, The College Board's goal is to create a scale so that all scores on the verbal and math portions of the SAT will have a mean of 500 and a standard deviation of 100. In 1994 the College Board revised the scale, and the actual mean on the math SAT for 1995 was 504 with a standard deviation of 110. What numbers would you use to label this normal curve?

e. Assuming Mary took the test in 1995, would her z-score increase or decrease? Why? Recalculate Mary's z-score using the new values.

Additional Problems

5-64. The College Board scales the scores with a maximum of 800 and a minimum of 200.

a. What is the corresponding range for the *z*-scores on the SAT's? (Assume a mean of 500 and standard deviation of 100.)

b. Simone just got back her SAT results and received 800 on the verbal section. Does this imply that she did not skip any questions and answered all of the questions correctly? Explain.

5-65. A set of data has a mean of 92 and a standard deviation of 8.

a. How many standard deviations above the mean is a raw score of 104?

b. If a student has a raw score of 110, what is his *z*-score?

c. What *z*-score corresponds to a raw score of 88?

d. What raw score corresponds to a *z*-score of –1.75?

5-66. Consider two complete sets of test data:

Test A	57	76	79	95	67	89	73	68	84
Test B	29	93	89	95	72	75	90	87	52

a. Find the mean (μ) and standard deviation (σ) of each set.

b. Who would have more reason to rejoice — the person who earned a 95 on Test A or on Test B? Explain your answer using your results from part (a).

5-67. Aimee has been tutoring her brother, Logan, for the past two weeks. As an incentive, she promised to take him out to dinner if he earns an A on his next test. (His teacher has announced that to earn an A, one's *z*-score must be ≥ 1.5.) The day came and Logan scored 42 out of 60. If the class average was 30 with a standard deviation of 8, did Logan earn his dinner? Explain.

5-68. What does it mean if someone has a negative *z*-score?

5-69. On a test with a mean of 66 and a standard deviation of 15, a *z*-score of 1 corresponds to what raw score?

5.3.1 How can I make predictions?

Normal Density Function

5-70. The Endowment Headwear Company makes printed visors as gifts for women running in charity marathons. Marathons in different cities have different numbers of runners. Endowment Headwear needs to know how many hats of each size to print for each event. The company collected the data below for women's head circumference by measuring 40 randomly-selected volunteers at charity events.

head circumference (cm)							
55.5	53.7	53.7	53.4	53.1	55.2	53.9	53.8
53.1	52.2	54.8	54.5	53.2	52.3	55.3	53.2
51.9	53.1	53.1	52.2	51.2	55.4	53.3	51.4
52.6	53.7	52.7	52.8	51.9	54.3	55.4	53.7
53.0	52.7	53.0	54.6	52.5	52.9	53.1	51.9

checksum 2133.3

a. What are the mean head circumference and standard deviation for these women?

b. On your calculator, make a histogram of the distribution of women's head circumferences and sketch it. Use an interval from 50cm to 57cm with a bin width of 1.

c. Hat size is determined by the whole-number portion of a woman's head circumference. For example a woman with a head measuring 55.8cm would need a size 55 hat. According to your histogram, how many women need a size 52 hat? Use the $\boxed{\text{TRACE}}$ button on your calculator to find the height of a histogram bar. On your calculator, press $\boxed{\text{WINDOW}}$ and set Ymin = –4 so the trace is easier to see.

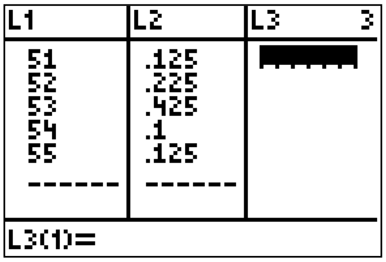
d. What *percent* of women have a hat size less than 54?

5-71. Questions about percentages, like those in parts (b) and (c) of problem 5-70 above, are easier to answer if you make a table of **relative frequencies**. Relative frequency is the **proportion** of hats: the *percent* of sizes out of the total 40 hats, written as a decimal.

a. Copy and complete the following table.

women's hat sizes		
size interval	frequency (# of hats)	relative frequency (proportion of hats)
50–51	0	0
51–52	5	0.125
52–53	9	0.225
53–54	17	
54–55		
55–56		
56–57		

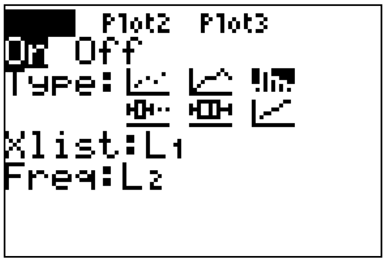
b. Now we will make a **relative frequency histogram**. In List1 on your calculator, enter the bins. In List2, enter the relative frequencies. Your calculator screen should look like this:



L1	L2	L3 3
51	.125	
52	.225	
53	.425	
54	.1	
55	.125	
------	------	

L3(1)=

On your calculator, make a relative frequency histogram from these two lists. Tell your calculator the relative frequency is in List2 by entering Freq:L2 on the STAT PLOT screen as shown at right.



Plot2 Plot3
On Off
Type:
Xlist:L1
Freq:L2

Sketch the relative frequency histogram. Label the top of each bar with its relative frequency. What do you notice about the shape of the relative frequency histogram?

c. Relative frequency histograms make percentages easier to visualize. Use the histogram to compute the percent of hats between size 52 and 55.

d. What percent of hats are below size 56? Show a computation.

5-72. The 40 women in the sample recorded their race times in various charity 5K races in the table below.

race time (min)							
22.2	22.6	24.9	23.5	22.8	23.3	23.1	21.6
21.3	22.9	25.7	23.3	23.3	22.5	24.4	22.7
24.1	23	22.5	23.2	24.7	24.4	23.3	23.5
23.1	22.5	22.3	22.6	23.6	23.3	23.3	23.4
23.0	23.1	24.5	23.9	20.6	23.5	22.8	24.4

checksum 928.7

a. Find the mean and standard deviation of the races times to four decimal places. Justify your choice of standard deviation.

b. Create a relative frequency histogram with your calculator, and sketch it. Use an interval from 19 to 27 with a bin width of 1. Label the top of each bar with its relative frequency.

c. Use the relative frequencies on your histogram to calculate the percentage of racers that had a time faster than 22 minutes. Remember, smaller times are faster.

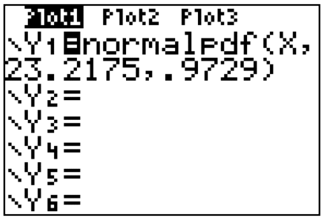
d. What percentage of racers completed a race between 22 and 25 minutes?

5-73. By creating a mathematical model of data we can describe the data to others without giving them a list of all the data, and we can make predictions based on the data. You may have already seen mathematical models of data when you drew a line of best fit as a model for scattered data. From the line of best fit, you were able to describe the association in the data to others, and you were able to make predictions from the model.

Much of the real data that is encountered in science, business, and industry can be modeled with a bell-shaped curve, called a **normal probability density function**. The mathematical formula is complex, but your calculator can draw it very easily once you determine the mean and standard deviation of the data.

Let's model the women's 5K times with a normal probability density function. If it is not already on your calculator screen, recreate the relative frequency histogram from part (a) of problem 5-72. Use an interval from 19 to 27 with a bin width of 1.

To model your data with a normal probability density function, press Y= and enter 2nd DISTR normalpdf/(X*,mean,standard deviation*). Your screen should look like the one at right.

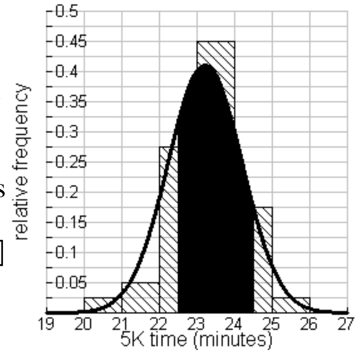


```
Plot1 Plot2 Plot3
\Y1=normalpdf(X,
23.2175,.9729)
\Y2=
\Y3=
\Y4=
\Y5=
\Y6=
```

Press GRAPH to display your histogram and the bell-shaped normal curve that is a model of the data. How well does your model represent the data? What are the strengths and weaknesses of your model?

5-74. If we want to use the relative frequency histogram to find the percent of women who have run times between 22.5 and 24.5 minutes, we would add up the bars between 22.5 and 24.5. But since our histogram is not drawn conveniently with those bins, we would have to redraw the histogram. Our model of the data comes to the rescue!

a. The height of the bars between 22.5 and 24.5 can be modeled with the area under the normal curve between 22.5 and 24.5, as shaded black in the diagram at right.



To find the area between 22.5 and 24.5, which is the proportion (percent) of women that fall between 22.5 and 24.5 minutes, enter $\boxed{\text{2nd}}$ [DISTR] normalcdf(22.5, 24.5,23.2175, 0.9729) on your calculator. What percent of women have running times between 22.5 and 24.5 minutes?

b. Your model represents the percentages of women that have various run times. But it does much more than that—since the sample of 40 randomly selected women represents the whole population, your model can tell you about percentages in the whole population. In general, what percentage of women have race times between 20 and 25 minutes? In general, enter $\boxed{\text{2nd}}$ [DISTR] normalcdf(*lower limit, upper limit, mean, standard deviation*) to have your calculator find the proportion for the interval.

c. Even though the fastest (smallest) time in our sample of 40 women was 20.6 minutes, the model that we chose—a normal probability density function—starts at negative infinity and goes all the way to positive infinity. What percentage of all women in the population run faster than 26 minutes according to your model? Since your calculator does not have an infinity key, instead enter -10^{99} for the lower limit.

d. Make a prediction with your model for the percentage of women that fall below the mean running time. Use -10^{99} for the lower limit, and the mean, 23.2175, for the upper limit. Does your answer make sense?

5-75. The Endowment 5K Race for Charity is coming up. The Endowment Headwear Company expects 775 racers. Based on a model for hat size distributions for all women, they will need to order hats for the event.

a. On your calculator, recreate the relative frequency histogram for the 40 women's hat sizes in problem 5-70. Use an interval from 50cm to 57cm with a bin width of 1.

b. Find the mean and standard deviation of the hat sizes. On your calculator, make a normal model of the data by pressing $\boxed{\text{Y=}}$ and entering $\boxed{\text{2nd}}$ [DISTR] normalpdf(X,*mean*,*standard deviation*). Sketch the histogram and the model.

c. According to your normal model, what percentage of women wear a size 51 hat? Shade this proportion on a new sketch of the model, and calculate the proportion using normalcdf on your calculator.

d. How many size 51 hats should Endowment order for the anticipated 775 racers at the Endowment For Charity 5K race?

e. Use the model to predict how many of the racers at the Endowment For Charity 5K race are expected to have a hat size below 51. Sketch these proportions on a new sketch of the model.

f. How many racers would we expect to have a hat size over size 56? Between 51 and 56?

g. Does your answer to part (f) make sense when compared to the answers from part (e)?

h. What percentage of racers does your normal model predict have hat sizes between negative infinity and infinity? Does your model make a sensible prediction?

Additional Problems

5-76. North City High School has served the following number of lunches since the beginning of the school year.

Number of lunches sold per day							
584	695	618	675	632	678	595	689
738	677	630	755	745	660	613	594
774	640	671	698	576	785	723	
721	603	636	663	665	640	671	
652	666	703	606	661	605	774	
checksum 24711							

a. What are the mean and standard deviation? What is the five-number summary of the distribution?

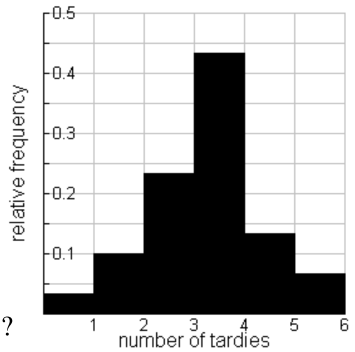
b. On your calculator, make a relative frequency histogram of the number of lunches served. Use an interval from 560 to 800 lunches, with a bin width of 40 lunches. Sketch the histogram and label the height of each of the bins.

c. Describe the distribution. Make sure you consider the center, shape, spread, and outliers.

d. Using your histogram, determine the percent of days that fewer than 600 lunches were sold.

e. Using the histogram, estimate the percent of days that between 600 and 700 lunches were sold.

5-77. Some students at North City High are abusing the privilege of being allowed to leave campus for lunch. The number of tardies to Mrs. Greene's period after lunch is too high. In the last 30 school days, she recorded the number of tardies shown in the histogram at right.



a. Describe the distribution of tardies by estimating the height of each bin.

b. On how many days were 3 or more students tardy?

c. On what percent of the days were no students tardy?

5-78. We are going to fit a normal model to the number of tardies in problem 5-77. The data Mrs. Greene gathered for 30 days is shown in the table below.

number of tardies per day					
2	4	3	3	4	2
1	3	3	4	3	2
2	3	1	2	3	3
3	0	2	3	2	5
5	3	3	3	4	1
checksum 82					

a. On your calculator, recreate the relative frequency histogram for the number of tardy students in problem 5-77.

b. Find the mean and standard deviation of the number of tardies.

c. On your calculator, make a normal model of the data by pressing o and entering 2nd [DISTR] normalpdf(X, *mean*, *standard deviation*). Sketch the model with the histogram.

d. According to your normal model, what percentage of days were 4 people tardy? Shade this proportion on a new sketch of the model, and calculate the proportion using normalcdf on your calculator.

e. Assume that the last 30 days in Mrs. Greene's class were representative of the 180 days in the whole school year. *According to your model*, how many days this year can Mrs. Greene expect 4 or more tardies? Sketch these days on a new sketch of the model.

METHODS AND MEANINGS

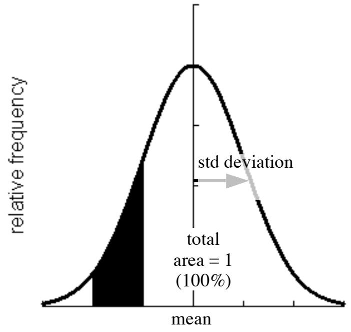
MATH NOTES

Normal Density Function

Much data in science, business, and industry may be represented by a bell-shaped curve. In order to be able to work mathematically with that data—describing it to others, making predictions—an equation called a normal density function is fitted to the data. This is very similar to how a line of best fit is used to describe and make predictions of data on a scatterplot.

The normal density function stretches to infinity in both directions. The area under the normal curve can be thought of as modeling the bars on a relative frequency histogram. However, instead of drawing all the bars, we just draw the curve that represents the top of all the bars. Like bars on a relative frequency histogram, the area under the normal curve (shaded in the diagram below) represents the portion of the population under study within that interval.

The total area under the curve is 1, representing 100% of the population under study. The mean of the population is at the peak of the normal curve, resulting in a portion of 50% of the population below the mean and 50% above the mean. The width of the normal curve is determined by the standard deviation (the variability) of the population under study: the more variability in the data, the wider (and shorter) the normal curve is.



5.3.2 How well did I do?

Percentiles

5-79. The ACT is a nationally-administered test that colleges use to help make admissions decisions for potential students. Nationwide, the scores are normally distributed, with a mean score of 21 (out of a possible 36) and a standard deviation of 4.7.

a. Sketch the normal model of the scores of all test takers on your calculator. An appropriate value for the maximum of the relative frequency axis is Ymax = 0.1.

b. Adèle scored 25 on the ACT. Let's explore how well she did. Remember your normal curve is a model for the bars of a histogram. Shade the "bars" (the area under your normal curve) for all the scores below Adèle's score. What percent of scores are below Adèle's score? Round to the nearest whole number.

c. Since 80% of test-takers scored below Adèle, we say that Adèle scored in the 80th **percentile**. What was Rémy's percentile? He scored 16 on the ACT.

d. Antoinette scored 21 on the ACT. *Without a calculator*, make a sketch of the scores below Antoinette on a normal curve and indicate what percentile Antoinette fell in.

5-80. Percentiles are not just for normal models. They can be used for any set of data to report the percent of scores that fell below a given score. The following data are the test scores on Mrs. Abraha's Chapter 3 test in geometry.

Chapter 3 Geometry Test Scores						
75	72	91	90	83	60	89
86	17	71	58	86	81	86
94	94	79	60	53	89	42
89	56	57	93	93	94	76
80	80	94	75	92	69	75

checksum 2679

a. Make a histogram of the data on your calculator and sketch it. Use a bin width of 5.

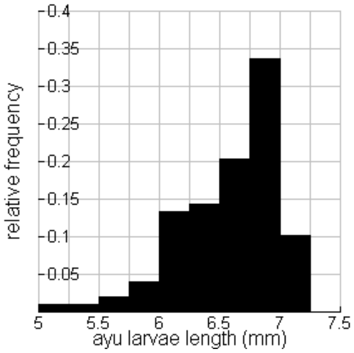
b. Lateefa is using a normal model, but her partner Farid thinks that it is not a good idea. What advice would you give this team about using a normal model?

c. Lateefa scored 86 on the test, and Farid scored 92. What percentile did they each score in?

d. What score is at the 25th percentile? 75th percentile?

e. Use 1-Var Stats on your calculator to find the five-number summary (min, Q1, median, Q3, max) for this data. How does your answer compare to part (d) above?

5-81. The March 2004 issue of *Ichthyological Research* reported the relative frequencies for the length of 98 larvae of the ayu sweetfish were as follows:



a. Estimate the 90th percentile for the length of ayu larvae.

b. Would you recommend making a normal model for the data to answer part (a)? Why or why not?

c. Estimate the 50th percentile. What is another name for the 50th percentile?

5-82. Rachna's physics class is going out to the football field to launch rockets today. The rocket that Rachna was assigned has historically had a mean launch distance of 74m with a standard deviation of 26m. A rocket's launch distance can be modeled with a normal curve.

a. Make a graph of the distribution of Rachna's rocket launch distances on your calculator, and sketch it. An appropriate value for the maximum of the relative frequency axis is Ymax = 0.02.

b. From your graph, visually estimate between what two distances the rocket lands 90% of the time? Shade this portion of your sketch.

c. Use your calculator to check your estimate. How close to 90% did you come?

d. The middle 90% (or 95% or 99%) of the data is an important computation in statistics. It tells us what "typical" data might look like without considering the small or large extremes at either end. Statistical computations will reveal that the lower limit of the middle 90% of normally distributed data is at z-score of -1.6449, and the upper limit is at z-score of 1.6449. Determine the middle 90% of distances the rocket will land more accurately than you did in part (b).

5-83. Rachna and her sister, Rakhi, were both in the same physics class. They made a bet with each other on who would launch their rocket farther. The loser would have to wash the family dishes for a month! Unfortunately they were assigned very different style rockets so it was difficult to compare. Rakhi's style of rocket had a mean launch distance of 30m with a standard deviation of 6m.

a. Rachna's rocket went 66.74m, while Rakhi's went 28.17m. Use percentiles to determine who had to wash the dishes for a month.

b. Use another method for determining who has to wash the dishes.

Additional Problems

5-84. In response to a judging controversy during the 2002 Winter Olympics, a new scoring system for ice dancing was used starting in 2006. The new system uses a "grade of execution" (GOE) as part of the overall score. The GOE goes from –3 to 3, and can be modeled with a normal density function with mean of 0 and standard deviation of 1. Isabella and Tony scored a 2 on the GOE. What percentile are they in? What percentage of ice dancers scored higher than they?

5-85. Due to natural variability in manufacturing, a 12-ounce can of soda does not usually hold exactly 12 ounces of soda. A can is permitted to hold a little more or a little less. The specifications for a soda-filling machine are that it needs to fill each can with a mean of 12 ounces of soda with a standard deviation of 0.33 ounces. Filling machines can be modeled with a normal density function.

a. Use your calculator to create a graph of a normal distribution using normalpdf. Sketch the graph. An appropriate value for the maximum of the relative frequency axis is Ymax = 1.5.

b. How often do you actually get a 12oz can of soda containing more than 12oz?

c. What percent of cans contain between 11.5 and 12.5 ounces of soda? Sketch these bottles on your diagram in part (a).

5-86. According to the *National Health Statistics Report*, the average height of adult women in the U.S. is 63.8 inches with a standard deviation of 2.7 inches. Heights can be modeled with a normal density function.

a. What percent of women are under 4ft-11in tall?

b. In North City High School's class of 324 senior students, how many girls would you expect to be shorter than 4ft-11in?

c. How many senior girls do you expect to be taller than 6ft at North City High?

5-87. Parking meters in a beach town cost 25¢ for 15 minutes. A normal model with mean \$10 and standard deviation of \$2 can be used to model the amount of money one parking meter makes on a busy summer day

a. Make a graph of the distribution of the money made by one meter on your calculator, and sketch it. An appropriate value for the maximum of the relative frequency axis is Ymax = 0.02.

b. From your graph, visually estimate between what two amounts of money the meter earns 90% of the time. Shade this portion of your sketch.

c. Use your calculator to check your estimate. How close to 90% did you come?

d. Statistical computations reveal that the lower limit of the middle 90% of the data is at z-score of -1.6449, and the upper limit is at z-score of 1.6449. Determine the middle 90% of the amount of money a single meter makes more accurately than you did in part (b).

UNIT 6 Data Collection

6.1.1 Samples and populations: what's the difference?

Population Parameters and Sample Statistics

The amount of information available to people continues to grow over time. Likewise, so does the demand for more current information. Statistical research is needed to supply this demand with unbiased information on a wide variety of topics so that the best possible conclusions can be made. Much of the information we obtain for statistical research begins as questions about populations of people or things.

6-1. Discuss the following questions with your team. Do not try to answer them, but instead provide an explanation of a realistic way to collect the information, and what kind of calculations would be necessary in order to answer each question.

a. Is there a relationship between the amount of physical activity a person gets and their perceived level of stress?

b. What is the mean 2009 SAT math score for the state of Arizona? Is it higher than the corresponding mean score for West Virginia?

c. What percentage of high school students would be willing to donate $10 or an hour of time to help a local food bank?

d. What is the mean weight of backpacks carried on college campuses? Is it different than the mean weight of backpacks carried on high school campuses?

e. Would wearing neckties increase standardized test scores for boys in middle school?

f. What is the average number of absences for the freshman class at your school?

g. What proportion of refurbished cell phones are defective?

h. How much cholesterol is in a chicken egg?

i. Do the seniors at your school do less homework than the sophomores?

MATH NOTES

METHODS AND MEANINGS

Population Parameters and Sample Statistics

Data collected from the entire group (the **population**) is called a **census**. Numerical summaries, such as mean and median, computed from a population are called **parameters**. A **sample** is a subgroup of a population and its numerical summaries are called **statistics**.

Examples:

If the members of our class were our **population** then a subgroup consisting of every fifth member of our class would be a **sample** and the mean test score of the sample would be a **statistic**.

If the members of our class were our **population** then the class test scores would be a **census** and the class mean test score would be a **parameter**.

6-2. Why sample? You must sample if there are too many items to measure, people to survey, things to count, or when taking a census is too expensive or time consuming. For example: What is the public approval rating of the Governor? What is the average commute time to work for U.S. employees? Also, the act of data collection in some cases ruins or destroys the item being measured. Examples: What is the bursting pressure of a 2 liter plastic bottle? How much vitamin A is in a carrot? What is the mean tensile strength of elevator cables?

When information is readily available from all of the people or objects of interest, there is no need to rely on just a sample. A census can be taken. Examples: What is the mean high school GPA of students admitted to Chapman University? What percentage of votes cast were for your U.S. congressman? What was the per capita spending on chewing gum last year?

Now look at your questions from problem 6-1 and decide if they can be answered in a practical way by using a census or a sample and whether the resulting calculations are statistics or parameters. Write your responses next to your answers for problem 6-1.

Additional Problems

6-3. Using the given population and your math class as a sample, come up with a research question that could be asked to determine parameters from the given population. Write each survey question and what you are hoping to show from each question. Examples are given for each population.

a. Given population: Students at my school

Example:

I would ask this survey question to my math class: *If the school library were open at 7:15 a.m., would you use it at least once per week?*

To answer this question about the students at my school: *If the library hours were extended, how many students at our school would take advantage of it?*

b. Given population: All high school math students in the U.S.

Example:

I would ask this survey question to my math class: *The main reason I'm taking this class is:*

1) It is a graduation requirement. *2) It is a college prerequisite.*
3) My parents are making me. *4) I'm interested in math.*
5) For some other reason.

To answer this question about all high school math students in the U.S.: *Why do high school students take math?*

c. Given population: U.S. teenagers

Example:

I would ask this survey question to my math class: *How far is it from where you live to school or work?*

To answer this question about U.S. teenagers: *What is the average commute distance to school for teenagers?*

6-4. Look again at each of the populations in problem 6-3. Does your math class accurately represent any of the populations in problem 6-3? How so, or why not?

6-5. For each question below, describe a realistic method to collect information to answer the question. Be sure to indicate if you would use a sample or census and if your results would be parameters or statistics.

a. What percentage of American League baseball players had a batting average above .300 this season?

b. How much pressure can be exerted on a chicken egg before it breaks?

c. How many hours of television does a high school student watch per day?

6.1.2 Is the question biased?

Detecting Bias in Survey Questions

When information needs to be gathered from human sources, often involving opinions or judgments, surveys must be used. Have you ever taken a survey or have you seen news reports that begin, “Surveys show that...”? Have you ever wondered how reporters draw the conclusions that they do? In this lesson, you will learn about surveys and evaluate whether they are likely to give biased results.

Have you heard statements like the ones below?

- “81% of public school students are not satisfied with the food provided by their schools.”
- “The President has an approval rating of 72%.”

How do news sources figure out this information? Do they ask every public school student? Do they ask every citizen? What questions do they ask?

Generally, to make claims such as the two above, someone has taken a survey. A survey is a type of sample. In this lesson, you will study surveys and consider how likely a survey is to give accurate results. You will start by considering how survey questions are asked and how their wording may influence the results.

As you think about survey questions, consider the following questions:

Is the question phrased in a way that will let me present the data easily?

Does the question influence the way people are likely to answer?

Is the question fair for everyone?

Can I influence the outcome by changing the way I ask the question?

6-6. Get a Lesson 6.1.2A Resource Page and answer the questions. Then your teacher will direct you in sharing your responses with the class.

a. Calculate the proportion of survey takers who responded in the given ways on Poll A and Poll B.

b. Were there any questions for which the responses were difficult to count?
How could these questions have been rewritten to make the responses clearer?

6-7. When answers to a question are influenced by the way the question is asked, the question is said to contain **bias**. The questions posed to your classmates in Polls A and B are reprinted below. There are descriptions in parts (a) through (f) below of some techniques unscrupulous pollsters often use to bias results.

Work with your team to decide which of the survey questions presented to your class use each of the techniques.

	Poll A	Poll B
1.	Do teenagers worry about their grades?	Do teenagers worry about getting poor grades?
2.	Do you support the Governor's education plan that ensures students will be more successful in school?	Do you support the Governor's education plan?
3.	Does violence in movies and video games affect young people?	Does the frequent occurrence of brutal violence in movies and video games have a negative affect on the young people exposed to them?
4.	Do you believe the current movie ratings system (G, PG, PG-13, R) is effective?	Do you believe the current movie ratings system (G, PG, PG-13, R) is effective?
5.	Should school districts spend more money on higher teacher salaries?	Should your math teacher be paid more?
6.	Moderate exercise is necessary to stay healthy. Do you exercise regularly?	Do you exercise regularly?

a. **Question Order:** Sometimes two questions are asked in an order such that the first question suggests an answer to the second. Which of the poll questions uses the "question order" technique? Why would you expect it to influence responses?

b. **Preface:** Some questions start with statements that can bias the result of the question that follows. Which of the survey questions presented to your class uses this technique? Why would you expect such statements to influence responses?

c. **Two Questions in One:** This technique involves asking two questions at once. Survey respondents may agree with one part and disagree with another part, but they are only allowed to give one answer. Which of the survey questions presented to your class uses this technique?

d. **Biased Wording:** By using pleasing or unpleasant words, the surveyor can influence results. Which of the survey questions presented to your class uses this technique?

Problem continues on next page. →

6-7. *Problem continued from previous page.*

e. **Desire To Please:** Research shows that many survey takers will answer in the ways that they perceive will please the surveyor. Which of the survey questions presented to your class are likely to be biased in this way? Are some more severely biased than others?

6-8. With your team, consider each of the survey questions below. Decide if any bias techniques are being used to influence the survey results. If no bias technique is being used, write "Fair question." Be prepared to explain your thinking to the class.

a. Jolly Juice has twice the Vitamin E of other brands. Which brand of juice is the healthiest?

b. Do you think that people who hurt defenseless animals should be punished?

c. Do you agree that Hal Poppington is the best man to be elected Mayor?

d. What is your favorite kind of juice?

6-9. You and your team members have been asked by the U.S. Department of Education to survey people about the President's new proposal to extend the school year from the current length of 180 days, to 200 days. A survey question might be, "Do you think students should attend school for 180 days as they do now, or for 200 days?"

Use the technique assigned by your teacher to rewrite the question to bias responses to be more favorable to the 200-day school year.

6-10. Survey questions can be either **open** or **closed**. Open questions allow free response, while closed questions have a limited number of possible responses. Open questions allow survey takers to express their ideas or opinions more accurately, but responses are generally difficult for surveyors to organize and analyze. Closed questions allow only a limited number of responses, which the researcher knows before conducting the survey; the advantage is that the resulting data is often easier to count and graph.

Label each of the following questions either "open" or "closed."

a. How often do you exercise?

A. Every day B. Once a week or more C. Less than once a week

b. What is your favorite way to exercise?

c. What is your favorite time of year?

d. In which country were you born?

6-11. For each question in problem 6-10 that you decided is open, give examples of answers that might be difficult to compare and quantify. For each question that you decided is closed, explain how the information you get from an answer may not be as accurate as it could be.

6-12. Through the next few class periods, you will work with a partner to participate in a survey project. Your teacher will assign you and your partner a survey question to work with. Work with your partner to analyze your question by answering the questions in parts (a) and (b) below.

a. Is your survey question open or closed?

b. It is possible for you to get too many different answers to your question to analyze your results well? If so, how could you reword your question so that your answers could be grouped and analyzed? Work with your partner to rewrite or change your question, if necessary. Ask your teacher to approve your new question.

Additional Problems

6-13. Consider each of the following survey questions. For each one, explain any bias you can find. If you think the question is unbiased (or fair), explain why.

a. Do you agree that it is important to make ending homelessness a high priority?

b. Which of the following factors is most important to address in order to slow global climate change?

A. Car emissions
B. Airplane emissions
C. Pollutants from private industry
D. Dependence on oil

c. How important is it that teacher salaries be raised?

6-14. Statistical results can be affected by psychological issues. Researchers who conduct experiments need to eliminate sources of bias for their statistical analysis to be meaningful. Discuss in your teams the potential causes of bias in the following samples.

a. Cola #2 did a taste test comparing itself to Cola #1. Participants were asked to pick their favorite drink—one labeled m and the other q. The majority of participants picked the drink labeled m, which was Cola #1.

b. A survey was conducted in the following manner: “The Bill of Rights guarantees the right to bear arms so that we can protect our families and our country. Recently, attempts have been made to enact stricter gun controls. Do you want these restrictions?”

c. Another survey was conducted in the following manner: “Last year over 15,000 people were murdered by handguns. That was 68% of all murders. Recently, attempts have been made to enact stricter gun controls. Do you want these restrictions?”

6.1.3 Does the sample represent the population?

Representative Random Samples

If you want to know what a bowl of soup tastes like, do you need to take a census or eat all of the soup in the cooking pot? Or, can you get a good idea of the taste by trying a small sample?

When conducting a survey or study, it is usually not possible to survey or study every person in the population you are interested in (for example, all the residents of the United States, all the students at your school, all teenage shoppers). Studying the whole population is often too time-consuming, costly, or impractical. But just like we can learn about the tastes of the soup by taking a sample, to learn about a population we can take samples.

But how can we take a sample that is representative of the whole population? For the soup, all we have to do is make sure it is well-stirred before tasting. But what about other populations that are not as easily stirred? That is the subject of today's investigation.

6-15. Some astronomers have long suspected that a belt of small objects exists between Mercury and the Sun. They call these Vulcanoid asteroids, named after the Roman god of fire. However, since the Sun is so bright we have never been able to directly observe these asteroids. Besides, an error in pointing the telescope can result in damage to the optics, and injury to the observer. Many other scientists do not believe Vulcanoid asteroids even exist. Suppose we sent a satellite with a special blackened optic to photograph these asteroids, should they exist. Can you imagine the excitement if the satellite began transmitting a picture of what appeared to be asteroids back to earth? It would be a huge moment in astrophysics and astronomy! Unfortunately, in just five seconds, the optics of the satellite were burned up and no one had a chance to record the video. If this scenario actually happened, there would still be many skeptics about the asteroids existence.

a. Even though you would only have seen the video the satellite sent back for five seconds, you could still estimate the size of a typical Vulcanoid asteroid. Model this scenario by looking at your "video" picture (Lesson 6.1.3 Resource Page) for five seconds and write down, without discussion, what size a typical asteroid is. Measure in millimeters.

b. Your teacher will give you instructions on how to make a histogram for the whole class. Write down a description of the histogram in context. Make sure you mention the center, shape, spread, and outliers.

6-16. Now suppose that the optics on the satellite were repairable and we were able to get a good photograph of the Vulcanoid asteroids. Then everybody would be a believer, and true scientific inquiry could begin!

a. The most important question remained: what size is a typical Vulcanoid asteroid? Without discussion with your team, pick ten "typical" asteroids and write down their numbers. Measure the diameter of each of these asteroids and calculate the mean diameter of a "typical" asteroid in millimaters.

b. Your teacher will again give you instructions about how to make a histogram of the "typical" asteroid size determined by each of your classmates. How does this new histogram compare to the histogram from the previous problem (center, shape, spread, and outliers)?

6-17. You are going to measure 10 asteroids again, but this time you are going to have your calculator randomly choose which ten asteroids to measure.

a. You will need to prepare your calculator for generating random numbers. Enter your seven-digit phone number as follows: *number* STO▶ MATH ***PRB rand***.

b. On your calculator, generate ten random numbers by entering MATH ***PRB randInt***(1,100,10) STO▶ ***L1***. List1 will now contain your ten random numbers between 1 and 100. Do not worry if you have repeated numbers. Each member of your team should generate their own random numbers.

Measure the ten asteroids that were randomly chosen. If you have repeated numbers, measure those asteroids twice. Calculate the mean diameter of your ten asteroids.

c. Your teacher will again give you instructions about how to make a histogram of the "typical" asteroid size determined by each of your classmates.

6-18. Compare the three histograms.

a. What do you notice about the center, shape, spread, and outliers?

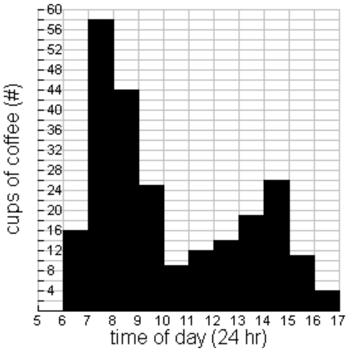
b. Which method produced the most consistent (least variable) results? Make a conjecture: which method produced the most accurate result for the mean size of a Vulcanoid asteroid? Why were you not very good at picking ten "typical" asteroids from the photo?

Additional Problems

6-19. Katelyn owns a small coffee cart in front of a downtown office building. Yesterday she kept track of how much coffee she sold each hour. She recorded the number of cups of coffee sold during a random sample of 30 hours, each hour during an 11-hour time period. (11:00a.m. is written 11:00 as usual, but 1:00p.m. is written as 13:00). Note that 6:00 and 16 indicates that 16 cups of coffee were sold from 6:00a.m. to 6:59a.m.

Time of day (24 hr)	6:00	7:00	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00
Cups of coffee (#)	16	58	44	25	9	12	14	19	26	11	4

Katelyn made the following histogram:



a. Using the data from the table and the histogram, estimate the five-number summary (min, Q1, median, Q3, max) without using your calculator.

b. Describe the distribution of coffee sales throughout the day.

6.1.4 Is there an easier way?

Cluster Samples

6-20. When you analyze results from your own survey, you will want to make claims about the thoughts or opinions of *all* of the students in your school. If you were to survey only students in the cooking club, for example, it might be hard to make claims about what all students think. Consider this idea as you think about the **samples** described below.

a. If you wanted to generalize the opinions of all students at your school, would it make sense to go to a bank and survey the people there? Why or why not?

b. If you wanted to generalize the opinions of all students at your school, would it make sense to ask all of your friends at school? Why or why not?

c. If you wanted to generalize the opinions of all students at your school, would it make sense to ask every third person at the school bus stop? Why or why not?

6-21. You learned in the last lesson that random sampling is much more representative of the population than intentionally trying to choose items to study. Randomization is a cornerstone of statistics.

Sometimes random sampling from the population is too impractical. Imagine if you wanted to survey voters in the United States about their preference of presidential candidate. Even if we had a way to randomly select 1000 names, how would we find those 1000 people to survey without it costing a fortune? Even at your much smaller school, if you selected a random sample of 50 people to survey it would be very time-consuming to find those 50 students and ask them survey questions.

Instead, statisticians sometimes use a **cluster sample**. They choose a "cluster," or group of people to survey that they believe represents the population well. A cluster sample is not ideal, but sometimes it is the only way to get a survey or observational study done.

What cluster of students might you survey at your school to represent all of the students at your school? Explain. Are there any reasons that this cluster might not be fully representative of all the students at your school?

6-22. What population do each of these samples represent? Write down the actual population for each of these sampling techniques.

Method of Sampling	Description of Actual Population
a. Call every hundredth name in the phone book.	People with phones who have their numbers listed; people who have a landline (in addition to a cell phone)
b. Call-in survey on TV	
c. Call people at home at 10 a.m.	
d. Ask every tenth person who leaves the mall.	
e. Ask people leaving the bank.	
f. An online "instant" poll or cell phone poll	
g. Mail questionnaires to people.	
h. Ask everyone on the school bus.	

6-23. In 1988, the steering committee of the Physicians Health Study Research Group released the results of a five-year experiment conducted on over 22,000 male physicians aged 40 to 84. The research on this sample suggested that the participants who took an aspirin every other day had a lower rate of heart attacks.

a. Can you legitimately conclude from this study that aspirin reduces the risk of heart attacks for all people? Why or why not?

b. Can you legitimately conclude from this study that aspirin reduces heart attacks for all men? Why or why not?

c. Can you legitimately conclude from this study that aspirin is linked to reduced heart attacks for all men aged 40 to 84? Why or why not?

d. Can you legitimately conclude from this study that aspirin reduces heart attacks for male physicians aged 40 to 84? Why or why not?

e. Can you legitimately conclude from this study that male physicians aged 40 to 84 should take aspirin? Why or why not?

6-24. With your team, come up with a survey question for students at your school. How will you choose a random sample of students to survey?

6-25. Suppose you were conducting a survey to determine what portion of voters in your town support a particular candidate for mayor. Consider each of the following methods for sampling the voting population of your town. State whether each is likely to produce a representative sample and explain your reasoning.

a. Call one number from each page of the phone book between noon and 2 p.m.

b. Survey each person leaving a local grocery store.

c. Survey each person leaving a local movie theater.

d. Walk around downtown and survey every fourth person you see.

e. Could you make a representative sample by surveying a few people from each of the situations described in parts (a) through (d) above? Explain.

6-26. On a square piece of paper provided by your teacher, write down an integer from 1 to 4. Tape it to the board to form a histogram.

6-27. Which numbers were picked most frequently? Discuss in your teams why this might have happened and how it applies to taking surveys.

Additional Problems

6-28. Bias can take many forms. Sometimes it is created unintentionally by conditions in an experiment. Other times it is more intentional. For each example below, comment on its possible bias.

a. In a TV commercial an interviewer asks people on the street to name their favorite radio station. All five that he asked claimed KCPM as their favorite.

b. A university campus wants to increase fees in order to build a new recreation hall. It surveys students to determine support. Survey booths are placed outside the university gym. The resulting survey showed overwhelming support.

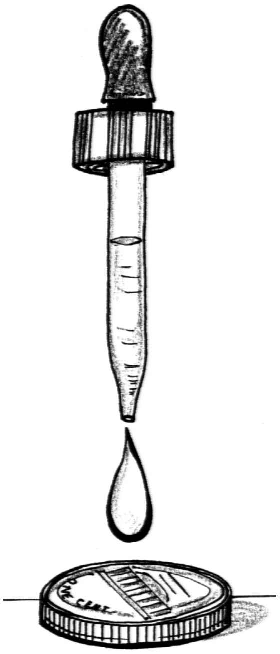
c. A study shows that many more accidents occur on I-95 during the day than at night. Is it therefore safer to drive at night?

d. An active citizens' group claimed that a nuclear facility was operating at below average safety standards and therefore should be shut down.

6-29. A music company wants to know the music preference of people in Cleveland. Their surveyor asks people who are walking out of a business office building, "What is your favorite type of music?" What kind of problems may arise that will not produce accurate results? (List as many problems as you can.)

6.1.5 How do my samples vary?

Sample-to-Sample Variability



6-30. How many drops of water do you think a penny can hold? Using your supplies, count the number of drops of water your penny can hold before it overflows. Record your results on a class dot plot.

6-31. Look at the dot plot generated by the data. Why is the data so spread out? Discuss as a class *exactly* how you conducted the study. Did everyone do it exactly the same way? What instructions could be given at the beginning of the observational study that would help make the results more consistent?

6-32. As a class, decide on uniform instructions and repeat the observational study.

6-33. Is the data more consistent this time around? Which study better answered the original question, "How many drops of water can a penny hold before it overflows?"

6-34. Make parallel box plots of the class data for the first and second trials. What conclusions can you draw about your different studies by comparing the box plots?

6-35. What are some of the reasons you still cannot answer the question "How many drops of water can a penny hold before it overflows?" exactly?

6-36. Find the class average, that is, the sample mean "number of drops." Do you believe it is the same as the population mean number of drops of water a penny can hold for the entire population of pennies and water drops? What are some of the reasons you cannot know for sure?

6-37. When you attempt to answer the question "How many drops of water can a penny hold before it overflows?" you are making a general statement about *all* pennies and *all* drops of water. However, the vast majority of pennies and drops of water in the world were not used in this study. Do your best to answer the question in plain language: What is the mean number of drops a penny can hold before it overflows?

MATH NOTES

METHODS AND MEANINGS

Sample-to-Sample Variability

The difference between the actual population parameter and a corresponding sample statistic can be due to **sample-to-sample variability**. Sample-to-sample variability exists because there is always some variation in the objects being measured or counted so no sample is an exact representation of the population it came from. Sample-to-sample variability is not the result of making mistakes in procedures or calculations, but rather it is a natural result of probability and variation.

To minimize the effect of sample-to-sample variability statisticians try to use the largest practical sample size. There are formulas and procedures for calculating appropriate sample sizes that you will learn about when you take a course in statistics. Statisticians can also help control variability in their data by using consistent testing procedures as your class did with their study of the pennies.

Additional Problems

6-38. Suppose the mean SAT score for a random sample of students is 960.

a. What does this tell you about the actual mean for all students who took the SAT?

b. Would another sample have the same mean?

c. For different samples, does the population mean change? Does the sample mean change?

d. How do sample and population means compare to each other?

6-39. Before distribution, batteries are tested to make sure that the mean life of the battery is within the acceptable limits set by the company. Give at least two reasons why the company only tests a sample and not all of the batteries.

6-40. Use sampling to estimate the mean number of words per page in your math book. Sample at least five pages. Record the page number and the number of words per page.

a. How did you select the pages to count?

b. What do you think would be the best way to select five pages for your sample?

6.2.1 Why can't studies determine cause and effect?

Association is not Causation

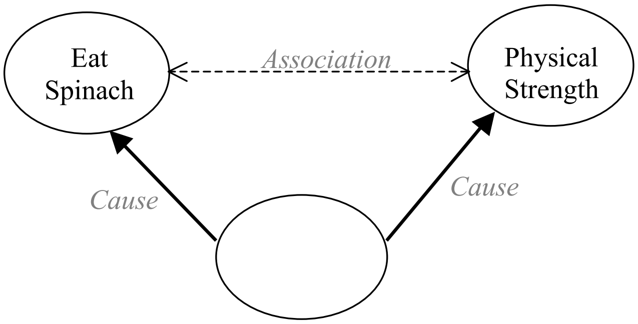
Another important distinction between how a statistical question can be answered is whether it measures the association or cause and effect between variables. As a consumer of information you need to be aware of the difference between finding association between variables and finding cause and effect.

6-41. A dietician studying the nutritional benefits of eating spinach reasoned that eating spinach would make a person physically stronger. The dietician surveyed a large sample of individuals and recorded their spinach consumption and their physical strength. After collecting and analyzing the data the dietician found the sample of spinach eaters to be significantly stronger than the sample of non-spinach eaters. Therefore the dietician found eating spinach is associated with (or linked to) physical strength and dietician *erroneously* concluded that the spinach was <u>causing</u> the observed increase in strength.

a. How is this possible? Where the dietician found spinach she found more average strength. Why is it incorrect to say spinach is the cause?

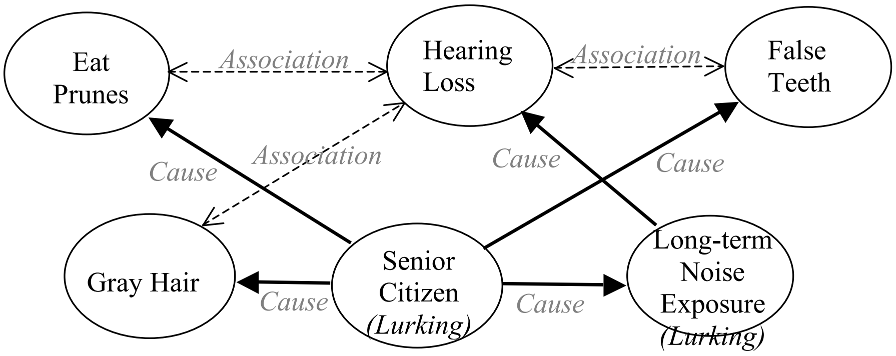
b. This finding could *correctly* be described as an <u>association</u> or link. The dietician found eating spinach is associated with (or linked to) physical strength.

Eating spinach and physical strength are variables in this observational study. Find another variable that could go in the black oval below that would actually cause someone to eat spinach and have physical strength. This is called a **lurking variable**.



6-42. Another researcher is looking for variables associated with concussions among athletes. There are many athletes in the world, some have had concussions and others not. An observational study could be performed and the researcher may *correctly* find a strong association between wearing helmets and having concussions. The researcher might *erroneously* conclude that helmets cause concussions in athletes. With your team, create a web of linked variables similar to the web in part (b) of problem 6-41, to show that the act of wearing a helmet is unlikely to cause a concussion. Be sure to label your arrows with the terms cause and association. What is/are your lurking variable(s)?

6-43. A web of linked variables can get complex and be impossible to unravel with an observational study. Consider a medical study focused on hearing loss. It may associate or link variables like eating prunes to having false teeth or gray hair to hearing loss as strongly as an actual cause like long-term noise exposure.



Here are some newspaper headlines from actual observational studies. Each of them touts a variable association or link and some even imply a cause and effect relationship. Your task is not to disprove the link, but to point out other lurking variables that may be part or all of the actual cause and effect. Determine at least one plausible lurking variable that could possibly explain causation. You may choose to draw a diagram like in the previous problems showing the relationships. Again, do not argue about the link expressed in the headline. Accept the association or link as true. Your task is to find the other variable(s) that could be the actual cause(s).

a. “Calcium in diet may cut risk for some cancers, study finds”

b. “Study: Family time declines as Web use booms”

c. “Chocolate is linked to depression”

d. “Study: Kids who were spanked have lower IQs”

e. “Study: Couples who cohabit more likely to divorce”

6-44. Now come up with your own original news headlines. The first sentence should contain a reasonable link between two variables. The second statement should be a clear misinterpretation of the link. Examples:

1st Statement: Facial Tissue linked to Colds and Flu.
2nd Statement: Surgeon General calls for a shift to handkerchiefs!

1st Statement: Bathing Suits tied to Sunburn.
2nd Statement: Dermatologist recommend: Swim Naked!

6-45. Make a poster of your news headlines from problem 6-44. You may want to include an illustration of the false cause and effect.

Additional Problems

6-46. Here are some more news headlines from real observational studies. Just as you did in problem 6-43, determine at least one plausible lurking variable that could explain the cause and effect. Remember, do not argue about the link expressed in the headline. Accept the association or link as true. Your task is to find the other variable(s) that could be the actual cause(s).

a. "Teens with own cars more likely to crash"

b. "Study connects hyperactivity, food additives"

c. "The graveyard shift may be aptly named. Working nights will soon be listed as a likely cancer cause"

d. "Daily meat diet tied to higher chance of early death"

6-47. Explain the difference between an association and causation. How could a researcher prove causation?

6.2.2 How can an experiment show cause and effect?

Experiments, Cause, and Effect

An **experiment** is based on a statistical question that requires some type of change or treatment be imposed on groups of subjects or objects so direct comparisons can be made. Experiments can be used to determine cause and effect relationships because the lurking variables, which plague observational studies, can be divided between the treatment groups using randomization and other techniques. Unfortunately experiments can be very expensive, time consuming, and in some cases involving humans or animals, even unethical.

6-48. Consider this question: "Does a new type of motor oil improve gas mileage?" Assuming this new motor oil was already available to the public, we could survey some car owners who use the oil and some who do not and compare their mean gas mileages.

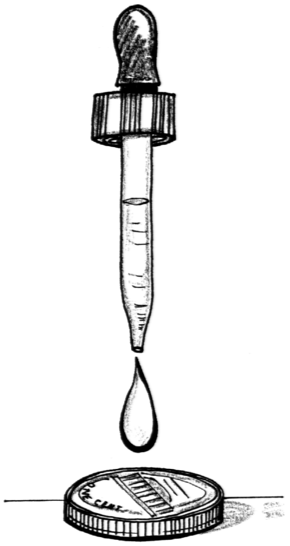
a. What problems might arise from using this method?

b. So what about the lurking variables like the age of the car, size of the engine, quality of gasoline, etc. in a motor oil experiment? If random assignment is used with a large sample size (30 is commonly considered large enough) for each group, then each group will likely have some hybrids and some non-hybrids, expensive and inexpensive gasoline, city and highway driving, etc. The lurking variables are more or less balanced so if there is an improvement in average mileage, the new oil is likely the cause. In an experiment, a control group gets no actual treatment. If people are involved they may get a dummy treatment called a **placebo** so the psychological effects of participating in an experiment are equally divided among the test groups.

Design an experiment to answer the original question.

6-49. The last time you worked with water drops on pennies it was an observational study. You did not seek to impose a change or treatment on the pennies. You were simply looking to see how much water some pennies in general circulation can hold. Now a treatment has been added to a random sample of pennies. Your teacher has cleaned a sample of pennies using a mild acid. The research question is "Do clean pennies hold a different amount of water than pennies in general circulation?"

You will be assigned a sample of pennies and a workstation to test your pennies. Use the same test procedures you used in Lesson 6.1.5. Make class dot plots for the clean pennies and the regular pennies. Use the information on the dot plots to create parallel box plots for a direct comparison of the class results.



6-50. As a class, discuss your results and write a brief conclusion.

6-51. Suppose your results from problems 6-49, and 6-50 were seriously biased by a lurking variable. In fact, they were. The effects of this lurking variable can be defeated by using proper experimental procedures. As a class, discuss improvements that could be made in the testing procedures.

6-52. Repeat the experiment using the revised testing procedures. Create the new dot plots and box plots. Write a conclusion to the question: "Do clean pennies hold a different amount of water than regular pennies?"

6-53. When determining if a question can, or should, be answered with an experiment, you must decide whether or not it would be ethical. In the previous lesson you were asked to find lurking variables that would be associated with the headline: "Study: Kids who were spanked have lower IQs". Certainly, there are plenty of lurking variables to cast doubt on cause and effect, but think about what it would take to do an experiment to answer this question. Design an experiment and point out why it is not ethical.

Additional Problems

6-54. Consider the question: "Does a traditional classroom SAT preparation course improve scores more than an on-line study course?"

Design an experiment that could help to answer this question. Refer to problem 6-48 for ideas.

6-55. Design experiments for the following statistical questions. If it would be unethical to conduct such an experiment, state why.

a. "Does listening to classical music during a math test improve scores?"

b. "Do seat belts save lives in car crashes?"

c. "Does vitamin C help prevent colds?"

MATH NOTES

METHODS AND MEANINGS

Observational Studies and Experiments

An **observational study** is based on a research question that seeks to measure the way things are without imposing any new factors (called **treatments**) on the subjects. A survey is one type of observational study. An **experiment** is based on a statistical question that requires some type of change or **treatment** be imposed on groups of subjects or objects. **Observational studies** are generally easier and less expensive to perform than **experiments** but seldom can determine cause and effect relationships because of the many hidden or **lurking variables**.

6.2.3 How can I answer the question?

Putting It Together

6-56. Look back at the questions in problem 6-1 (reprinted below). Categorize each question as a census or sample, and then decide whether an observational study or experiment should be carried out to answer the question. If an observational study is possible, explain how you would carry out the study. If the study asks for an association between variables, discuss the effects of at least two possible lurking variables. If an experiment would be suitable, outline an experimental design. If surveys are necessary, list a potential source of bias in the question(s), and a particular difficulty in getting a representative sample from the population.

a. Is there a relationship between the amount of physical activity a person gets and their perceived level of stress?

b. What is the mean of the 2009 SAT math scores for the state of Arizona? Is it higher than the corresponding mean score for West Virginia?

c. What percentage of high school students would be willing to donate $10 or an hour of time to help a local food bank?

d. What is the mean weight of backpacks carried on college campuses? Is it different than the mean weight of backpacks carried on high school campuses?

e. Would wearing neckties increase standardized test scores for boys in middle school?

f. What is the average number of absences for the freshman class at your school?

g. What proportion of refurbished cell phones are defective?

h. How much cholesterol is in a chicken egg?

i. Do the seniors at your school do less homework than the sophomores?

6-57. Answer some of the same questions from these actual newspaper headlines. If an observational study was done, explain how. Explain why an experiment was not possible. If the study shows an association between variables, discuss the effects of at least two possible lurking variables. If you believe an experiment was done, state so, and outline a possible experimental design. If surveys were necessary, list a potential source of bias in the question(s), and a potential difficulty in getting a representative sample from the population.

a. "Study sticks it to traditional back care. Acupuncture – real and fake – gets better results for pain than the usual treatments."

b. "MARITAL STRIFE A HEART WRECKER? Bad marriage can risk coronary disease risk, researchers say"

c. "Breastfeeding May Cut Breast Cancer Risk. Women with a family history of breast cancer who have ever breastfed reduce their risk of getting premenopausal breast cancer by nearly 60%, according to a new study."

d. "Study: Oral drug better than lotion to kill lice… A new study has found that in tough cases, an (new) oral medication kills the parasites more effectively than a prescription lotion applied to the scalp."

Unit 8 Sample-to-Sample Variability

8.1.1 How can I estimate complex probabilities?

Simulations of Probability

If you toss a coin ten times, what is the probability of having a run of three or more “heads” in a row?

If an airline “overbooks” a certain flight, what is the chance more passengers show up than the airplane has seats for?

When 67 people get cancer in the 250 homes in a small town, could that be due to chance alone, or is polluted well water a more likely explanation of the cluster of cancer cases?

When the mathematics becomes too complicated to figure out the theoretical probability of certain events, statisticians often use **simulations** instead. Simulations can also be used to check statistical computations, or if a study is too expensive, takes too much time, or is not ethical. A simulation is a model—often computer-based—that mimics a real-life situation.

All simulations require the use of random numbers. Random numbers have no pattern; they cannot be predicted in any way. Knowing a random number in no way allows you to predict the next random number.

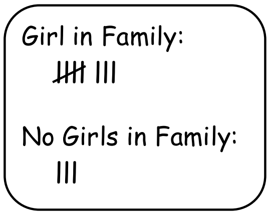
Complex simulations like modeling the weather, traffic patterns, cancer radiation therapy, or stock market swings require tens of billions of random numbers. For those simulations a large computer running for hours or even days is needed, but many simple simulations can be done with our TI-83+/TI-84+ graphing calculators.

8-1. Michael and Debbie want to have children and would love to have a girl. They are trying to calculate the chances of having a girl if they have children until they have a girl, or until they have four children, whichever comes first. The probability is a little complicated, and it would not be ethical (or practical) to do an observational study by randomly selecting 1000 couples and having them produce four children so Michael and Debbie can count the number of girls. And imagine how many years the study would take! So they will run a simulation by tossing a coin. Since a coin has a 50% chance of landing on “heads,” a coin can be used to model the real-life situation of the 50% probability of a girl being born.

Problem continues on next page. →

8-1. *Problem continued from previous page.*

a. Let “heads” represent a girl. Shake a coin in a cup and “pour” it out. If a boy is born, do not tally anything, and flip again. If a girl is born (“heads” comes up), put a tally mark by “girl in family” on your paper and *stop having children*. If four boys in a row are born, *stop having children* and tally “no girls.”



Girl in Family:
𝍸 III

No Girls in Family:
III

b. Repeat the simulation until you have modeled 25 families (25 trials). Each family consists of having children until either a girl is born, or four boys are born, and the result is tallied. You will have 25 tally marks when you are done. If you are finished early, continue tallying families until the rest of the class catches up.

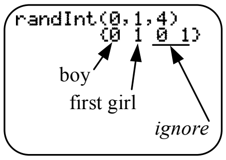
c. Combine your results with those of rest of the class. Then, according to your class’s simulation, what is the probability of having a girl if you have children until you have a girl, or until you have four children, whichever comes first?

8-2. According to a mathematical principal known as the **Law of Large Numbers**, the more times you run your simulation, the closer your result will approach the true theoretical value. Since tossing coins is tedious, we can use a computer to complete many more trials.

a. Start a new tally sheet.

b. You will use your calculator to randomly generate a family of four children, with “***0***” representing boys, and “***1***” representing girls. Since we stopped having children after our first girl, we can ignore all the digits in the family after the first girl. For example, “***0101***” would represent a family with one boy and one girl, and we would mark “Girl in Family” on the tally sheet.

Generate a family by entering $\boxed{\text{MATH}}$ ***PRB randInt(0,1,4)*** on your calculator. Does the family have a “girl” or “no girls?” Mark your tally sheet.



```
randInt(0,1,4)
       {0 1 0 1}
```

c. Press $\boxed{\text{ENTER}}$ to generate another family, and tally the result. Each member in your team can quickly generate 50 families. If you are finished early, continue tallying families until the rest of the class catches up.

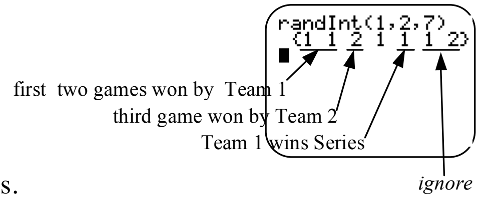
d. Combine your results with those of your classmates. Because you ran the simulation many more times, your class’s computer simulation will be closer to the true probability than the probability you found by tossing a coin. Use your class’s results to estimate the probability of having a girl if you have children until you have a girl, or until you have four children, whichever comes first.

Additional Problems

8-3. In the World Series of baseball, the first team to win four games wins the championship. The series might last four, five, six, or seven games. A fan who buys tickets would like to know how many games, on average, he can expect a championship series to last. Assume the two teams are equally matched, and set aside such potentially confounding factors as the advantage of playing at home. (If you've already completed some intermediate probability, you recognize this as the **expected value** for the number of games.)

a. Simulate a World Series by entering $\boxed{\text{MATH}}$ ***PRB randInt(1,2,7)*** on your calculator. Let a "***1***" represent Team 1 winning a game, and "***2***" represent Team 2 winning. You will simulate seven "games," but as soon as a team wins four games the World Series is over and you will ignore any additional games. See below for an example.

b. Record how many games it took to win the series. In the example above, it took 5 games.



```
randInt(1,2,7)
 {1 1 2 1 1 1 2}
```

c. Repeat the simulation at least 25 times.
Each time, record the number of games it took to win the World Series.
You do not care *which team* won, you only care *how long* the series took.

d. Based on your simulation, what is the average number of games played for a World Series?

8-4. Myriah hates doing the dishes, but her parents insist that she must help out and do them sometimes. Since Myriah wants to leave it to chance whether or not she will have to do the dishes, Myriah's mom proposes that they roll two dice. If the sum of the dice is 6 or less, Myriah will do the dishes. If the sum is 7 or more, one of her parents will do the dishes.

How often can Myriah expect to have to do the dishes? Run a simulation of rolling the two dice by entering $\boxed{\text{MATH}}$ ***PRB randInt(1,6,2)*** on your calculator, and record the sum. Press $\boxed{\text{ENTER}}$ to run the simulation again and record your sum. Run the simulation 30 times to determine how often Myriah can expect to do the dishes.

8.1.2 How many in a streak?

More Simulations of Probability

8-5. Twins Paolo and Paola are in the same math class. For homework, their teacher assigned them to flip a coin two hundred times. When they turned their assignments in, the teacher accused Paola of just making up results, rather than actually flipping a coin. Their results are below. Which homework is Paola's? Why?

```
THTTTTHHTHTHTTTTTHHTHTHHHHHHHHHTHTHTHHHTHHHTTHHTHTHHHHHTHH
THHTTHTHTHTTTTTTHHTHTHTHTTTHTTTTHHTTTTHTTTTHTTTTHHHTHTHHTT
THTTTHTTHHTHHHTHHHHTHHHTTHHHTHHTTTTTTTHTTTHTTTHTTTHTTTHTTH
HHHTHHHTTTHTTTTTTHHTHTTTHH
```

```
THTHTTTHTTTTHHTHTTTHTTHHHTHHTHTHTHTTTTHHTTHHTTHHHTHHHTTHHH
TTTHHHTHHHHTTTHTHTHHHHTHTTHHHTHHTHTTTHHTHHHTHHHHTTHTHHTHHH
TTTHTHHHTHHTTTHHHTTTTHHHTHTHHHHTHTTHHTTTHTHTHTTTHTHHTTHTTT
HTTTTHHHHTHTHHHTTHHTHHTHHT
```

8-6. Investigate how many streaks of five or more heads or tails in a row you would expect when you tossed a coin two hundred times.

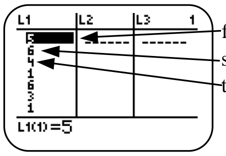
a. Since this is a complicated probability, a computer simulation will help. Model the coin toss by entering MATH ***PRB randInt(0,1,200)*** STO▸ 2nd L1 into your calculator. In List1 you will find the 200 coin tosses.

b. Use a tally sheet to count the number of streaks of five or more (it does not matter if it is a streak of heads or a streak of tails).

c. Share your results with your team. How did the teacher know that Poala made up her results?

8-7. Katelyn is going to babysit her nephew many times this summer. She had the great idea that one way to entertain him is to walk to McBurger's for a Kids Meal for lunch each time. The Kids Meal comes packed randomly with one of six possible action figures. Katelyn is worried that her nephew may be disappointed unless he gets all six action figures. Katelyn would like to know how many meals she can expect to buy for her nephew before getting all six figures.

a. Model the action figures with the digits 1 through 6. Help Katelyn simulate buying Kids Meals by entering MATH ***PRB randInt(1,6,200)*** STO▶ 2nd [L1]. In List1 you now have simulated Kids Meals, each with one of the action figures #1 to #6. How many meals did you have to buy in order to get all six action figures?



first meal: figure #5
second meal: figure #6
third meal: figure #4

b. Run your simulation 10 times. Keep track of how many Kid's Meals you have to buy each time to get all six action figures.

c. Combine your results with those of your team. On average, how many Kids Meals will you need to buy in order to get all six action figures?

d. Estimate a range on the number of meals you might need to buy. What's the most you needed to buy? The least? Do you think it possible that you may have to buy 50 meals to get all six action figures? 100 meals?

Additional Problems

8-8. Sports announcers frequently get excited when basketball players make several free throw shots in a row. They say things like “he’s on a hot streak tonight!” or “he’s really in the zone—what an amazing performance!”

a. Are these “hot” streaks really special, or are they just a natural run to be expected by probability? We will use simulation to determine what really is an unusually long streak of free throws, as opposed to a streak that is expected through normal play. Assume a basketball player has a 50% free throw average, and a typical game has 20 free throw attempts. Use your calculator to randomly generate 20 free throws, with “***0***” representing a missed shot, and “***1***” representing making the free throw. Enter MATH ***PRB randInt(0,1,20)*** STO▸ 2nd L1. In List1 you now have 20 free throws. How long was the longest streak of making free throws (what was the most number of “***1***”s in a row)?

b. Run the simulation 25 more times. Each time record the length of the longest streak.

c. How long would a streak have to be before you considered it unusual?

8-9. Myriah hates doing the dishes, but her parents insist that she must help out and do them sometimes. Since Myriah wants to leave it to chance whether or not she will have to do the dishes, Myriah’s mom proposes that they roll two dice. If the sum of the dice is 6 or less, Myriah will do the dishes. If the sum is 7 or more, one of her parents will do the dishes.

It’s bad enough when Myriah has to do the dishes. But Myriah really hates doing the dishes several days in a row! Run a simulation of rolling the two dice by entering MATH ***PRB randInt(1,6,2)*** on your calculator, and record the sum. Press ENTER to run the simulation again and record the sum. Repeat 90 times to simulate the ninety days in three months. How often can Myriah expect to have to do the dishes 3 or more days in a row during a month?

8-10. A never-ending game?

Jack and Jill are playing a game where Jack has 2 pennies and Jill has 4 pennies. A coin is tossed. If it lands on heads Jill has to give a penny to Jack. If it lands on tails, Jack gives a penny to Jill. The game is won when one of them has all of the pennies. They think that they can keep playing all day since the coin has an equal chance of landing on heads or tails, so they will just keep passing coins back and forth.

Simulate this situation on your calculator. Can they keep playing all day, or does one player have a better chance of winning the game?

8.1.3 How much do my samples vary?

Simulating Sampling Variability

You may have seen in Unit 6: Data Collection that if a sample is selected randomly and with care to avoid bias, we can be confident that the sample represents the whole population. Making an inference from a sample to the whole population is at the heart of what statistics is all about.

Your friend Ramien says that you have to have the latest, hottest π-Phone—everybody already has it! What proportion of teens really do have the π-Phone? We cannot possibly ask every teen in the U.S., but we can take a random sample of teens and calculate the proportion that have the phone.

If a random sample of 1000 teens finds that 250 have the π-Phone, it does not mean that exactly $\frac{1}{4}$ of all teens in the U.S. have the phone. The proportion of teens with the phone will naturally vary from sample to sample—some samples of teens will have more phones, some will have less.

If we knew how much the proportion naturally varied from sample-to-sample, we could establish a range of estimates for the proportion of phones in the teen population. This **margin of error** is frequently reported in statistical studies. You might read in a newspaper that the proportion of teens with a π-Phone is 23% with a margin of error of 3%. This means that statisticians believe from their small sample of teens that between 20% and 26% of all teens own the π-Phone.

8-11. What proportion of candy-coated chocolates are red? Since we cannot count every candy in the world, we can take a sample. We can assume a bag of candy-coated chocolates makes a reasonable sample that represents the whole population. We will investigate how much the proportion of red candies will naturally vary from bag to bag.

a. Calculate the proportion of red candies in your sample. Write your answer as a decimal and share it with the class.

b. Use your calculator to make a histogram of all the proportions of red candies your classmates found in each of their samples. A histogram that shows the results of taking many samples is a called a **sampling distribution**. Sketch the histogram.

c. What is the mean of the whole class's proportions? This gives an estimate of the proportion of red candies in the population.

d. You are interested in typical results, not the extreme ones. Out of all the proportions, you are interested in the middle 90%. We need to find an upper and lower bound for the class's proportions.

Your teacher will show you how to sort the whole class's proportions on your calculator. What proportion are about 5% of the samples greater than? What proportion are about 5% of the samples less than? Between what upper and lower bounds are the middle 90% of the red proportions?

e. Finally, predict the proportion of red candies in the whole population and give the margin of error.

Additional Problems

8-12. The Bright Idea Lighting Company wants to determine what proportion of LED flashlights that come off of its assembly line are defective. It takes many samples of 100 flashlights over the week and determines the proportion that are defective in each sample. The results of their tests follows:

Proportion of defective flashlights in 100 samples.				
0.09	0.05	0.05	0.05	0.09
0.07	0.06	0.10	0.07	0.09
0.08	0.09	0.05	0.05	0.08
0.07	0.07	0.08	0.08	0.09
0.10	0.08	0.11	0.08	0.09
0.09	0.13	0.08	0.09	0.09
0.07	0.06	0.06	0.08	0.08
0.06	0.10	0.11	0.09	0.09
0.05	0.07	0.05	0.09	0.10
0.04	0.07	0.09	0.08	0.06
0.08	0.09	0.06	0.06	0.07
0.08	0.08	0.08	0.09	0.09
0.10	0.07	0.09	0.08	0.12
0.12	0.09	0.06	0.05	0.02
0.11	0.05	0.05	0.05	0.06
0.08	0.10	0.08	0.10	0.06
0.06	0.07	0.08	0.11	0.09
0.09	0.08	0.07	0.07	0.05
0.10	0.07	0.11	0.06	0.06
0.10	0.08	0.10	0.10	0.07

(checksum 7.83)

a. What is the mean proportion of defective flashlights (as a percent)?

b. Use the technique of problem 8-11 to find the upper and lower 5% bounds of the sample-to-sample variability.

c. Predict the proportion of defective flashlights (in percent) in the whole population and give the margin of error.

8-13. A survey of 80 seniors at Algieres High School found that 65% had been accepted to a four-year university. A computer simulation was conducted to determine the sample-to-sample variability:

Proportion of students attending four-year universities. Simulation mean 0.65.				
0.66	0.63	0.73	0.63	0.69
0.65	0.68	0.58	0.73	0.54
0.65	0.65	0.66	0.62	0.68
0.51	0.59	0.71	0.63	0.64
0.59	0.68	0.67	0.64	0.59
0.64	0.67	0.71	0.59	0.61
0.75	0.60	0.68	0.63	0.72
0.74	0.60	0.61	0.67	0.74
0.61	0.67	0.56	0.58	0.62
0.71	0.73	0.65	0.67	0.63
0.64	0.63	0.59	0.71	0.70
0.57	0.62	0.57	0.62	0.71
0.56	0.67	0.65	0.62	0.56
0.77	0.76	0.67	0.63	0.62
0.52	0.73	0.69	0.62	0.69
0.62	0.61	0.63	0.54	0.67
0.63	0.68	0.65	0.59	0.65
0.66	0.65	0.70	0.61	0.65
0.65	0.71	0.66	0.76	0.62
0.68	0.62	0.59	0.67	0.60

(checksum 64.64)

a. Use the technique of problem 8-11 to find the upper and lower 5% bounds of the sample-to-sample variability.

b. Predict the proportion of the seniors at Algieres High that have been accepted to four-year universities and give the margin of error.

8.1.4 Should I take a larger sample?

Sampling Variability With Increased Sample Size

8-14. A new high school is built on the other side of town. Its school colors are blue and gold. How rude! *Your* school colors are blue and gold. The Student Council surveys a random sample of 25 students to see if students wish to change the school colors to maroon and gold. Only 20% of the students supported the change from their beloved blue and gold. We will investigate the margin of error of this survey.

a. We will use a computer simulation to estimate the sample-to-sample variability. Since 20% of the students support a change in colors, we will use the first 20 numbers (from 1 to 20) to represent a student who supports the new colors; the numbers from 21 to 100 represents one of the 80% of students who does not want to change school colors. Randomly choose 25 "students" from the "population" in which 20% support the change by entering MATH ***PRB randInt(1,100,25)*** STO▶ 2nd [L1]. Your 25 students are stored in List1. What proportion (percentage) of the 25 students in your sample supported the new colors? Write your answer as a decimal.

b. Repeat the simulation with 5 more trials and record the proportion that support the new colors each time. If you are finished early, continue with more simulations until the rest of the class catches up.

c. If you were to list all the proportions of your classmates, what do you suppose the mean of all the proportions would be close to? Why?

d. Combine your results with those of your classmates. What is the margin of error?

8-15. The Student Council thought it might be more helpful to survey more students, perhaps 100 students instead of 25.

a. Why would the Student Council want to survey more students?

b. Do you think that the margin of error for 100 students will be larger or smaller than for 25 students? Why?

c. Simulate choosing 100 students. MATH ***PRB randInt(1,100,100)*** STO▶ 2nd [L1]. Repeat the simulation five times. If you are finished early, continue with more simulations until the rest of the class catches up.

d. Combine the results with those of your classmates. What is the margin of error for 100 students?

e. Do you need to modify your conjectures in parts (a) and (b)? Explain your results.

Additional Problems

8-16. A consumer magazine randomly selects 250 of its readers and asks if their luggage was lost on their last airplane flight. Seven out of the 250 lost their luggage. The consumer magazine conducted a simulation to determine the sample-to-sample variability and concluded that about $3\% \pm 1.7\%$ of all passengers lose their luggage. If the magazine had surveyed 1000 readers instead, make a conjecture about what the number of passengers who lose their luggage might be?

8-17. Daylight Saving Time was adopted in the U.S. in 1918. During the oil crisis of the 1970s, the Department of Transportation found that daylight saving time decreased national energy usage by about 1 percent compared with standard time. Since then, energy use in the U.S. has changed and daylight saving time has been extended. In 2007 a group of researchers found that daylight saving time decreased national energy usage by 0.2% with a margin or error of 1.5%. If a 1% decrease in energy use will save the state of Indiana $9 million, what conclusion(s) can you draw from the 2007 study in relation to the state of Indiana?

8-18. A copy machine company advertises that its copiers will make at least 25,000 copies before requiring maintenance. A consumer research group tested the claim by collecting data from users of the particular copy machine in 30 various regions of the country. The mean for each of the 30 regions is listed in the table below.

24928	24574	24652	24758	24691
24893	25024	24767	24791	24609
25249	24914	24895	24656	24883
24551	24928	25025	24798	25041
24782	25020	24618	24904	24764
24705	24889	24656	24600	24735

checksum 744300

Use the technique of problem 8-11 to find the upper and lower 5% bounds of the sample-to-sample variability and predict the number of copies that can be made before a machine requires maintenance. Do you think the consumer research group will support the company's claim?

MATH NOTES

METHODS AND MEANINGS

Sample-to-Sample Variability

To learn about a population, we take samples when studying the whole population is too time-consuming, tedious, or impractical. Even if we are very careful to avoid bias in our sample so that it represents the whole population, the statistics (measurements) of a sample vary naturally from sample to sample. To make an inference about the population we need to quantify this **sample-to-sample variability**. With this knowledge, we can then state a margin of error for our prediction of the true parameters of the whole population.

In Lesson 8.1.3 you took many samples by hand to estimate the sample-to-sample variability. Since taking many samples is often not practical, computer simulation, like you conducted in this lesson, is another way to estimate the sample-to-sample variability. You simulated conducting many surveys of 25 students to see how much they varied.

8.2.1 Can I make a decision?

Statistical Test Using Sampling Variability

8-19. The principal of Algieres High School cancelled the Winter Formal dance because he believed that students preferred an all-school trip to an amusement park during spring break. But he told the Student Council that if they could convince him that more than 50% of the school preferred the Winter Formal he would reinstate it. The Student Council randomly surveyed 25 students; 60% of them preferred to keep the dance. Even though a majority in the sample wanted the dance, the principal was not convinced. He said that due to natural sample-to-sample variability, the true proportion of dance supporters might not actually be more than 50%.

This kind of investigation into sample-to-sample variability is called a hypothesis test: a claim has been made about the population (more than 50% support the dance), and we take a sample to test whether there is convincing evidence for the claim.

a. We will need to investigate the sample-to-sample variability of a survey that concludes 60% of students support the dance. Our computer simulation will generate random numbers from 1 to 100. The first 60 numbers (from 1 to 60) will represent a student who supports the dance, while the numbers from 61 to 100 represent a student who does not support the dance. Randomly choose 25 "students" from the "population" in which 60% support the dance by pressing [MATH] ***PRB randInt(1,100,25)*** [STO▶] [2nd] [L1]. Your 25 students are stored in List1. Due to natural sample-to-sample variability the proportion of students who support the dance in your random sample is probably not exactly 0.60. What proportion of students in your sample supported the dance?

b. Repeat the simulation with 5 more trials and record the proportion of dance supporters each time.

c. If you were to combine your proportions with those of your classmates, what do you suppose the mean of all the samples would be close to? Why?

d. Combine the results of your samples with those of your classmates. What are the lower and upper 5% bounds of your sample-to-sample variability?

e. Predict the proportion of the all the students at Algieres High that support the dance and give the margin of error.

8-20. The principal will consider your prediction in part (e) of problem 8-19 for the proportion of all students at the high school that support the dance. Does the principal have convincing evidence that more than 50% of the students support the dance?

Additional Problems

8-21. The principal at Algieres High School believes that over 20% of students are text-messaging at least once a week during class—obviously he believes this is an impediment to effective learning. Students claim that the actual percentage is much lower and that stricter rules are not necessary. Mrs. Rahil secretly observed her homeroom class very carefully for a week; only 13% of students actually text-messaged during class. Assuming her homeroom is representative of the whole school, Mrs. Rahil's class did a computer simulation and determined a margin of error of 10%. Is it plausible that the principal is correct and 20% of the whole school is text-messaging?

8-22. Students at Algieres High think they are safer drivers than average teenagers. In a random sample of 120 students at Algieres High, 18 students said that they got into a car accident when they were 16 years old.

a. What percent of 16-year-olds in this sample got into a car accident?

b. You wish to do a computer simulation of the sample-to-sample variability. Explain exactly what you would type into your calculator to do this simulation. After you did the simulation once, what would the numbers in your calculator screen represent?

c. Billy ran the simulation 100 times and obtained the following proportions of 16-year-olds that got into a car accident.

Proportion of 16-year-old students in a car accident. Simulation mean = 0.15				
0.19	0.14	0.13	0.12	0.20
0.13	0.17	0.15	0.16	0.14
0.12	0.13	0.14	0.12	0.17
0.11	0.13	0.16	0.18	0.20
0.10	0.17	0.15	0.17	0.19
0.14	0.19	0.11	0.15	0.10
0.12	0.14	0.19	0.09	0.13
0.18	0.20	0.14	0.11	0.17
0.20	0.15	0.20	0.18	0.15
0.11	0.16	0.12	0.20	0.14
0.17	0.10	0.18	0.17	0.19
0.18	0.11	0.11	0.14	0.13
0.15	0.13	0.18	0.17	0.28
0.14	0.15	0.13	0.11	0.13
0.11	0.11	0.15	0.15	0.12
0.20	0.15	0.22	0.17	0.11
0.15	0.18	0.12	0.26	0.15
0.13	0.11	0.13	0.15	0.16
0.14	0.15	0.23	0.15	0.08
0.19	0.15	0.15	0.18	0.10

(checksum 15.17)

d. Consider the lower and upper 5% bounds to determine the margin of error for the proportion of students who got into an accident at Algieres High.

e. According to the U.S. Census Bureau, 21% of 16-year-olds nationwide get into a car accident. But in the sample at Algiers High, only 15% got into a car accident. Can students at Algiers make the claim that they are safer drivers?

MATH NOTES

METHODS AND MEANINGS

Margin of Error

Using a statistic from a representative sample to make an inference about a population parameter is what the bulk of statistics is all about. Due to natural sample-to-sample variability you cannot say for sure what the true value for the population is based on a statistic. But you do have confidence that the true value lies within an interval called the **margin of error**.

To compute a margin of error, a computer simulation is used to estimate the sample-to-sample variability. Then the middle 90% (or 95% or 99%) of the data are used as the upper and lower bounds of our estimate. By sorting the simulated samples from lowest to highest, and determining where the lowest 5% and upper 5% of the data begin, we can determine the margin of error. The margin of error is half this range. Data from 100 simulated samples from a population in which 48% support the president might look like this:

43% 44% 44% 45% 45% 45%…(88 more samples)…50% 51% 51% 52% 52% 54%

The lowest five samples (lowest 5% of 100 samples) is at 45% and the upper five samples are at 51%. So the range is 6%, and we would report that 48% of the voters support the president with a margin of error of ±3%.

8.2.2 Did my experiment show results?

Variability in Experimental Results

Today you will use what you know about sample-to-sample variability to determine if two results in an experiment are truly different.

8-23. The female red-eyed poison-dart frog visits numerous bromeliads ("air plants") where pools of water have collected in the leaves. In each pool, she lays a single egg that grows into a tadpole. The tadpoles feed on mosquito larvae in the pools, but locals have been killing the mosquito larvae. Environmentalists are concerned because they are unsure whether or not tadpoles will adapt and eat some other kind of food source.

Janelle led a team of environmental scientists to the Costa Rican rain forest. They tagged 100 female tadpoles and determined the number that grew to adulthood. Mosquito larvae were placed in 50 of bromeliads, and the other 50 bromeliads were treated so that mosquitoes could not live there.

36 of the 50 tadpoles (72%) with mosquitoes survived, while 29 of the 50 (58%) non-mosquito tadpoles survived. The difference in the proportion that survived is $0.72 - 0.58 = 0.14$. Investigate whether a difference in the proportion of tadpoles that survived of 0.14 can be explained by natural sample-to-sample variability, or if there is a true difference between the two groups.

a. First explore the sample-to-sample variability using a computer simulation. When there are two samples, we can roughly estimate the sample-to-sample variability by combining the samples. For our model, there are 100 tadpoles, numbered 1 to 100. There were $36 + 29 = 65$ tadpoles that survived, so the numbers 1 to 65 will represent a tadpole that survived. The numbers 66 to 100 will represent a tadpole that died. Take a sample of 100 tadpoles with your calculator: [MATH] ***PRB randInt(1,100,100)*** [STO▸] [2nd] [L1]. What proportion survived? What proportion did not survive? What is the *difference* between these two proportions (proportion survived minus proportion that did not survive)?

b. The difference of proportions in your random sample that survived is probably higher or lower than 0.14, or even negative, due to natural sample-to-sample variability. What does a positive difference mean in the context of this problem? What does a negative difference mean?

Problem continues on next page. →

8-23. *Problem continued from previous page.*

c. If we take repeated random samples, we can estimate the sample-to-sample variability, and thus the margin of error. Here are the differences in the proportion that survived (proportion survived minus proportion that did not survive) in 100 trials of the simulation in part (a):

Proportion survived minus proportion did not survive. Simulation mean = 0.14				
0.10	0.33	0.13	0.13	0.10
0.05	0.10	0.02	–0.05	0.14
0.21	0.12	0.22	0.22	0.17
0.13	0.24	–0.07	0.23	0.14
0.16	0.30	0.12	0.21	0.11
0.17	0.19	0.14	0.13	0.09
0.05	0.10	0.23	0.29	0.30
0.26	0.08	0.07	0.30	0.05
0.12	–0.04	0.16	0.03	0.13
0.26	0.16	0.08	0.14	0.07
0.27	0.19	–0.01	0.19	0.16
0.06	0.17	0.18	0.06	0.13
0.15	0.11	0.07	0.05	0.15
0.18	0.24	–0.02	0.10	0.16
0.19	0.18	0.36	0.20	0.06
0.22	0.17	0.16	0.26	–0.04
0.13	0.05	0.05	0.15	0.19
0.13	0.18	0.10	0.13	0.10
0.16	0.21	0.17	0.09	0.03
0.13	0.13	0.12	0.15	0.23

(checksum 14.1)

Now find an upper and lower bound on the proportions. What proportion are about 5% of the samples greater than? What proportion are about 5% of the samples less than? What upper and lower bounds are the middle 90% of the proportions between?

d. You have been looking at the sample-to-sample variability when you simulated the differences in proportions between those that survived and those that did not survive to be 0.14. Predict the true difference in the proportion of those that survived and those that did not survive and give the margin of error.

e. Is a difference of zero a plausible result considering your margin of error? What does a difference of zero mean in the context of this problem?

f. Are you convinced that there is a true difference in the tadpoles that ate mosquitoes and those that did not?

Additional Problems

8-24. Students in Miss Hampton's science class tested the effectiveness of detergent in getting dishes clean. They created a gooey paste of hard-to-clean foods (spaghetti sauce, mustard, mashed potatoes, and grape jelly) and smeared 250 clean dinner plates with an exact amount of the food paste. They weighed each plate and randomly placed them into commercial dishwashers. Half the dishwashers had detergent in them, and half had only clean water. After cleaning the dishes, they weighed each plate to determine the portion of food paste that remained. 84% of the food was removed from dishes cleaned with detergent, while only 72% of the food was removed from dishes cleaned without detergent. Using the steps below, explore whether detergent really helps dishes get cleaner than just plain water.

a. What is the difference in the proportions (detergent minus plain water)? Express your answer as a decimal.

b. Mrs. Hampton's class ran a computer simulation and determined the sample-to-sample variability of the *difference* between the proportion of food removed by the detergent compared to plain water. They concluded that the *difference* in the true proportion of food removed was 0.12 ± 0.085.

Is a difference of zero a plausible result considering their margin of error? What does a difference of zero mean in the context of this problem?

c. Are you convinced that there is a true difference between cleaning with detergent and cleaning with plain water?

8-25. Olivia loves playing putt-putt golf. In putt-putt golf you do not swing at a golf ball, but rather you only putt the golf ball (tap the ball with a club so that it rolls into a hole). Olivia experimented with her new club to determine whether or not she played better with it. Each time before she putted a golf ball, Olivia flipped a coin to determine whether she would use her new club or the old one. She experimented on 80 putts. With the new club, Olivia made 25% of the 40 putts, while with the old one, she made only 15% of the 40 putts. Using the steps below, help Olivia decide whether or not her new golf club is really better, or if this difference can be explained by sample-to-sample variability.

a. What is the difference in the proportions (new club minus old club)? Express your answer as a decimal.

b. Olivia needs a computer simulation to determine the sample-to-sample variability:

   i. Out of the 80 putts, how many went into the hole?

   ii. For our computer model, there are 80 putts. What will the numbers 1 to 16 represent? What will the numbers 17 to 80 represent?

   iii. Conduct a simulation of 80 putts by entering `MATH` ***PRB randInt(1,80,80)*** `STO▸` `2nd` [L1] into your calculator. What proportion of the putts in simulation went into the hole? What proportion did not go into the hole? What is the *difference* in the proportion (proportion that went into the hole minus proportion that did not go in)?

Olivia ran the simulation 50 times and calculated the *difference* in the proportion of putts that went into the hole and those that did not go into the hole for each simulation. From her results she predicted the true *difference* in proportion of all her putts was 0.10 ± 0.146.

c. Is a difference of zero a plausible result considering your margin of error? What does a difference of zero mean in the context of this problem?

d. Are you convinced that there is a true difference between the new club and the old club?

MATH NOTES

METHODS AND MEANINGS

Statistical Tests

You have looked at two different types of **statistical tests**. In Lesson 8.2.1, you took a sample and compared the sample statistic to a **claim** about the population. Specifically, you compared the survey results of 60% to the principal's claim that the true value of the population was 50% or less. By simulating the sample-to-sample variability you created a margin of error. If the claim was within the margin of error, you concluded that the claim was plausible. If the claim was not within the margin of error, you concluded that the claim was not plausible.

In the experiment with frog tadpoles you performed a different kind of statistical test. You had two samples, and you calculated the *difference* between those two samples. Then you looked at the sample-to-sample variability of the differences between the two samples. A margin of error was created for the true difference between the two populations. If a difference of zero was within that margin of error, you concluded that it was plausible that there was no difference between the two populations; you could *not* conclude that the two populations were in fact different. On the other hand, if a difference of zero was not within the margin of error, you could conclude that it was plausible that there was actually a true difference between the two populations.

8.2.3 Should I reject for poor quality?

Quality Control

8-26. When companies accept items for shipment or distribution, they often do not test every single item for quality. In many processes, testing every item would be far too time-consuming and expensive. Or, consider the quality control testing of 2-liter soda bottles for bursting strength. Sometimes testing even destroys the item! If they tested every item there would be nothing left to ship. Instead, quality control testing is done on a small sample of items. If the sample fails in quality, the entire batch is thrown out or recycled.

Your task is to assure quality control for the P.C.I. (Probability Cubes Incorporated) dice company. The quality of their product is critical for the success of P.C.I. We plan to take 10 dice out of each case and test to see if they fall within company specifications. If the dice fail to fall within the acceptable range, the entire case will be thrown out.

To determine the acceptable range for well-functioning dice, we will use a computer simulation.

a. Each person in your team should simulate the rolling of a perfectly manufactured set of ten dice. Enter [MATH] ***PRB randInt(1,6,10)*** [STO▸] [2nd] [L1]. In List1 you now have the simulated rolls of ten dice. Use ***1-Var Stats*** to find the mean of the ten rolls and record your mean.

b. Each team member should repeat the simulation five times. Record the mean each time.

c. Your teacher will tell you how to combine your mean with those of all your classmates. If you are done early and waiting for your classmates, continue to repeat the simulation and record the mean.

d. Combine your results with those of your classmates. Make a histogram of the sampling distribution. What is the mean of the sampling distribution?

e. Find the upper 5% and lower 5% bounds as you did for previous lessons. Compare your bounds with the rest of the class.

8-27. We will consider the bounds you found in problem 8-26 the upper and lower bounds of a normal manufacturing process of ten dice. A set of ten dice that does not fall within this range will be considered defective and the entire case will be thrown out.

a. Your teacher will have one or more samples of ten dice from P.C.I., each from a different case that was manufactured early this morning. Roll the sample of ten dice once and calculate the mean roll. Did your batch fall within the quality control bounds? If not, P.C.I. will need to discard the entire case they came from.

b. Repeat for another sample of dice if you have one. Was this sample of dice within your quality control bounds?

8-28. How many cases of dice were rejected by the quality control engineers in your class? What percent of the cases were rejected?

Additional Problems

8-29. Due to natural variability in manufacturing, a 12-ounce can of soda does not usually hold exactly 12 ounces of soda. A can is permitted to hold a little more or a little less. The specifications for the soda-filling machine are that it needs to fill each can with 12 ± 0.25 ounces of soda. If a can of soda is filled with 11.97 ounces of soda, is the filling machine operating within specifications?

8-30. A company makes metal plates on which they put a special coating. Flaws appear in the finish of these metal plates and the company wants to establish a quality control system. A machine scans each plate after it is finished and reports the number of flaws. The data they collected for 50 samples are shown below.

7	10	9	3	13	7	5	8	8	10
13	9	21	10	6	8	3	12	7	11
5	10	6	13	3	2	7	4	11	8
1	7	5	2	0	11	3	4	3	1
4	14	7	12	6	1	10	6	2	12

checksum 361

Determine quality control bounds for the company. Careful! Should you reject plates that have too few flaws?

9.2.6 How can I use it? What's the connection?

Extension: Analyzing a Game

The activities in this section review several big topics you have studied so far. Work with your team to decide which combination of tools you will need for each problem. As you work together, think about which skills and tools you are comfortable using and which ones you need more practice with.

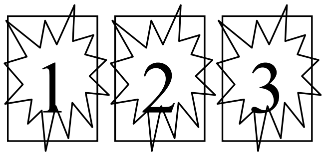
As you work on this activity, keep in mind the following questions:

What mathematical concepts did you use to solve this problem?

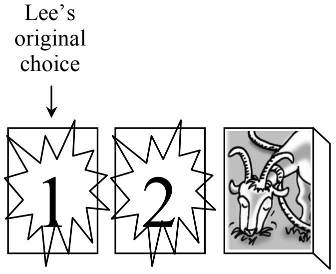
What do you still want to know more about?

What connections did you find?

9-82. THE MONTY HALL PROBLEM



Wow! Your best friend, Lee, has been selected as a contestant in the popular "Pick-A-Door" game show. The game show host, Monty, has shown Lee three doors and, because he knows what is behind each door, has assured her that behind one of the doors lies a new car! However, behind each of the other two doors is a goat.



"Which door do you pick?" Monty asks.

"I pick Door #1," Lee replies confidently.

"Okay. Now, before I show you what is behind Door #1, let me show you what is behind Door #3. It is a goat! Now, would you like to change your mind and choose Door #2 instead?" Monty asks.

What should Lee do? Should she stay with Door #1 or should she switch to Door #2? Does she have a better chance of winning if she switches, or does it not matter? Discuss this situation with the class and make sure you provide reasons for your statements.

9-83. Now test your prediction from problem 9-82 by simulating this game with a partner using either a computer or a programmable calculator. If no technology is available, collect data by playing the game with a partner as described below.

Choose one person to be the contestant and one person to be the game show host. As you play, carefully record information about whether the contestant switches doors and whether the contestant wins. Play as many times as you can in the time allotted, but be sure to record at least 10 results from switching and 10 results from not switching. Be ready to report your findings with the class.

If playing this game without technology, the host should:

- Secretly choose the winning door. Make sure that the contestant has no way of knowing which door has been selected.
- Ask the contestant to choose a door.
- “Open” one of the remaining two doors that does not have the winning prize.
- Ask the contestant if he or she wants to change his or her door.
- Show if the contestant has won the car and record the results.

9-84. Examine the data the class collected in problem 9-83.

a. What does this data tell you? What should Lee do in problem 9-82 to maximize her chance of winning?

b. Your teammate, Kaye, is confused. “Why does it matter? At the end, there are only two doors left. Isn’t there a 50-50 chance that I will select the winning door?” Explain to Kaye why switching is better.

c. Gerald asks, “What if there are 4 doors? If Monty now reveals two doors with a goat, is it still better to switch?” What do you think? Analyze this problem and answer Gerald’s question.